Journal of Digital Science

ISSN 2686-8296

Volume 3 Issue 2 **December 2021**

CONTENTS

Coronavirus Genome Sequence Similarity and Protein Sequence Classification

Partha Mukherjee[1[0000-0001-9244-8600]],
Youakim Badr[1[000-0002-8976-7894]],
Srusthi N. Karvekar[1[0000-0001-8913-3368]],
Shanmugapriya Viswanathan[1[0000-0003-4459-2274]]

[1]The Pennsylvania State University, Great Valley, Malvern, PA-19335, USA

https://doi.org/10.33847/2686-8296.3.2_1

Received 26.10.2021/Revised 01.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. The world currently is going through a serious pandemic due to the coronavirus disease (COVID-19). In this study, we investigate the gene structure similarity of coronavirus genomes isolated from COVID-19 patients, Severe Acute Respiratory Syndrome (SARS) patients and bats genes. We also explore the extent of similarity between their genome structures to find if the new coronavirus is similar to either of the other genome structures. Our experimental results show that there is 82.42% similarity between the CoV-2 genome structure and the bat genome structure. Moreover, we have used a bidirectional Gated Recurrent Unit (GRU) model as the deep learning technique and an improved variant of Recurrent Neural networks (i.e., Bidirectional Long Short Term Memory model) to classify the protein families of these genomes to isolate the prominent protein family accession. The accuracy of Gated Recurrent Unit (GRU) is 98% for labeled protein sequences against the protein families. By comparing the performance of the Gated Recurrent Unit (GRU) model with the Bidirectional Long Short Term Memory (Bi-LSTM) model results, we found that the GRU model is 1.6% more accurate than the Bi-LSTM model for our multiclass protein classification problem. Our experimental results would be further support medical research purposes in targeting the protein family similarity to better understand the coronavirus genomic structure.

Keywords: Coronavirus Disease of 2019 (COVID-19), Severe Acute Respiratory Syndrome (SARS), Genome Structure, Basic Local Alignment Search Tool (BLAST), Gated Recurrent Unit (GRU), Protein Family Accession.

1. Introduction

1.1. Background

In December 2019, an acute respiratory disease caused by a newly identified beta(β)-coronavirus, occurred in Wuhan, China and received attention all over the world. Initially, this coronavirus was named as the 2019-novel coronavirus (2019-nCoV) on January 12, 2020 by the World Health Organization (WHO). Coronavirus Study Group (CSG) of the international committee proposed to name the new coronavirus as SARS-CoV-2 whereas WHO officially named the disease as coronavirus disease 2019 (COVID-19) on February 11, 2020 [1]. On 7 January 2020, Chinese scientists rapidly isolated a SARS-CoV-2 from a patient to isolate the genome sequence of the SARS-CoV-2 [2].

SARS-CoV is a member of the Coronoviridae family of enveloped, positive stranded RNA viruses, which have a broad host range. Coronavirus infections in rodents, cats, pigs and cattle can be responsible for enteric diseases or cause respiratory diseases in people, cattle and birds. Twenty-three putative proteins,

including four major structural proteins; nucleocapsid (N), spike (S), membrane (M), and small envelope (E) encode from the 27-32 kb genomes of coronavirus. Out of these, the spike protein on the viral surface facilitates the entry and viral attachment to the host cell. In addition, variations of S protein among strains of coronavirus are responsible for host range and tissue tropism. However, the S, M, and N mature proteins all contribute to generate the host immune response as seen in transmissible gastroenteritis coronavirus, infectious bronchitis virus, pig respiratory coronavirus, and mouse hepatitis virus [3].

Among the several coronaviruses that are pathogenic to humans, most are associated with mild clinical symptoms, with two notable exceptions: 1) severe acute respiratory syndrome (SARS) coronavirus (SARS-COV); and 2) Middle East Respiratory Syndrome (MERS) coronavirus (MERS-COV). SARS-COV was a novel beta-coronavirus that emerged in Guangdong, southern China in November, 2002. It resulted in more than 8000 human infections and 774 deaths in 37 countries during 2002–03. MERS-COV was first detected in Saudi Arabia in 2012 and was responsible for 2494 laboratory-confirmed cases of infection and 858 fatalities since September 2012, that included 38 deaths following an outbreak into South Korea [3].

1.2. Origin and Transformation

The SARS-CoV-2 is a β-coronavirus, which is enveloped non-segmented positive-sense RNA virus. Coronaviruses (CoV) are divided into four genres, including alpha, beta, gamma, and delta. Alpha-CoV and beta-CoV can infect mammals, while gamma-CoV and delta-CoV tend to infect birds. Previously, six CoVs have been identified as human-susceptible virus strains, among which alpha-CoVs such as: 1) HCoV-229E, and 2) HCoV-NL63; and beta-CoVs such as: 1) HCoV-HKU1, and 2) HCoV-OC43 with low pathogenicity cause mild respiratory symptoms similar to a common cold respectively. In HCoV-XXXX, the pattern "XXXX" represent the strain names such as 229E, HKU1, OC43 etc. The other known beta-CoVs, SARS-CoV and MERS-CoV lead to severe and potentially fatal respiratory tract infections [4]. The genome related terminologies are listed in the glossary shown in Table 1 below.

In this study, we perform genome structure analysis and compare the level of similarity among genome structures of coronavirus genomes isolated from a COVID-19 patient, SARS patient and a bat.

Further we classify the protein structure of all the genome structure into different protein family accession to identify prominent protein family accession present in all the three viruses. The protein accession number is an identifier for family of proteins with similar functions and structure. Each family has an accession number named by pfam / uniport.

The remaining sections of this study is arranged as follows: In section 2, we investigate the related works performed as part of medical and technological studies in alignment with the COVID-19 virus. In section 3, we present and conduct the genome structure analysis in order to find prominent protein families in the viruses. We preprocess the dataset and perform data exploratory analysis in section 4. In section 5, we discuss the experimental results and their significance with respect to answering our research questions. Finally, we conclude our work and identify future work in section 6.

Table 1. Glossary of terms

GLOSSARY
Potive-Stranded RNA viruses Some viruses, such as coronaviruses, carry their genetic material as RNA rather than the more typical DNA-based genomes. Positive stranded RNA (also called plus-stranded) indicates that the single stranded RNA genome is of the same sense as coding messenger RNA
Beta (β)-Coronavirus It is in the subfamily Orthocoronavirinae in the family Coronaviridae, of the order Nidovirales. They are enveloped, positive-sense, single-stranded RNA viruses of zoonotic origin.
SARS-CoV The strain of coronavirus causing global outbreak of contagious and sometimes fatal respiratory illness which appeared in China in 2002.
MERS-CoV The coronavirus causing respiratory illness which appeared in the Arabian Peninsula in 2012.
CODON A sequence of 3 bases in a m-RNA strand which corresponds to a amino acid. (Eg: UAA, AUG, etc.)
Open Reading Frames (ORF) A continuous stretch of codons that begins with a start codon (usually AUG) and ends at a stop codon (usually UAA, UAG or UGA)
Non-Structural Proteins (NSP) A protein encoded by a virus but that is not part of the viral particle.
HCoV Refers to Human coronavirus
MetaTranscriptomic Refers to the science that studies gene expression of microbes within natural environments

2. Literature Review

Coronaviruses either leads to respiratory diseases or enteric infections in various animal species. These viruses can also cause hepatic and neurological diseases. Human coronaviruses, identified in the 1960s, are responsible for up to 30% of respiratory infections with prototypes HCoV-OC43 and HCoV-229E. The three serotypes of coronavirus have three main classes identified by the phylogenetic analysis [5].

A large number of studies have proved that the pathogen of COVID-19 is a novel coronavirus till date. The COVID-19 pathogen has a linear single-stranded positive-strand RNA genome of about 30 kb that belong to the Beta (β) Coronavirus genus and Sarbecovirus subgenus [6].

The complete genome isolated from a COVID-19 patient who worked in the Wuhan sea-food market, has one strain of size 29.9 kb [5]. SARS-CoV and MERS-CoV have positive-sense RNA genomes of 27.9 kb and 30.1 kb, respectively [7]. It has been stated that the genome of coronavirus contains a variable number, up to six to eleven, of Open Reading Frame (ORFS). Most of the RNAs located in the first ORF translates two polyproteins and encodes 16 Non-Structural Proteins (NSP). This forms two-thirds of the viral RNA. The remaining ORFs encode accessory proteins, that

interfere with the host's innate immune response and structural proteins. The four essential structural proteins include spike (S) glycoprotein, small envelope (E) protein, matrix (M) protein, and nucleocapsid (N) protein [2].

Deep Meta-Transcriptomic sequencing on genome isolated from Wuhan patient, performed by Wu et al. [8] contained 16 predicted NSPs. When compared with the SARS-Cov and MERS-Cov, this genome sequence was closer to the SARS-like bat coronavirus. At the protein level, there are no amino acid substitutions that occurred in NSP7, NSP13, with the exception in NSP2, NSP3, spike protein. NSP(XX) are the names isoforms of the NSPs, (XX is the integer number) where the naming is done chronologically and the numbers (i.e., XX) may not have any significance on protein structures [8]. Another recent research suggested that the mutation in NSP2 and NSP3 play a role in capability of infection, and differentiation mechanism of SARS-CoV-2 [9]. These studies influence the researchers for exploring the difference in the host and transmission between the SARS-CoV-2 and SARS-CoV and consequently opens avenue of further exploration to find the potential therapeutic targets [5].

Ruan et al. [3] analyzed the genotypes of COVID-19 in different patients from several provinces and found that SARS-CoV-2 had been mutated in different patients in China. Tang et al. [10] conducted a population genetic analyses of 103 SARS-CoV-2 genomes and classified out two prevalent evolvement types of SARS-CoV-2, L type (~ 70%) and S type (~ 30%). The L type strains are derived from S type and are more contagious in nature. Thus, this novel coronavirus needs to be monitored closely to inspect this pandemic.

Full-genome sequence analysis of COVID-19 revealed that SARS-CoV-2 belongs to beta (β) coronavirus, but it is divergent from SARS-CoV and MERS-CoV that caused epidemics in the past. The COVID-19 along with the Bat-SARS-like coronavirus forms a distinct lineage within the subgenus of the Sarbecovirus [11].

One of the long-term problems in molecular biology is to understand the relationship between an amino acid sequence and the protein function. This insight will result in useful scientific implications. As discussed in Bileschi et al [12], the classification of 17000 protein families is not feasible with conventional machine learning techniques like SVM, Decision Trees and ensemble models. Conventional machine learning models have limitations such as the requirement of substitution matrices, hard tuned scoring functions, feature engineering and sequence alignment to mention a few. By leveraging advanced Artificial Intelligence (AI) technology and especially Deep Learning methods, we overcome these constraints and directly predict the protein functional annotations from the data. Deep Learning algorithms efficiently co-locates sequences from unseen families with more accuracy [13]. The Deep Learning model learns the relationship between unaligned amino acid sequences and their functional annotations across protein families of the Pfam database [14]. Deep sequencing technology facilitates parallel processing of numerous distinct genome fragments and identify millions of base pairs within a few hours. The genomic analysis using deep sequencing determines: a) the structure and location of genes, b) regulatory elements, c) non-coding RNAs, and predict the gene functions [15]. For these reasons, we propose Deep Learning based model GRU to predict the protein family accession number with the protein sequence and compare it with the performance of Bi-LSTM method [16-18] using the same dataset. Our model captures the patterns between the sequence formation and protein function that cannot be easily identified. To the best of our knowledge, the GRU technique in this study has not been explored and thus can be used for creating the protein profile of the newly emerged coronavirus.

3. Research in Context

In the previous section, we described the genomic characteristics of COVID-19 with similarities and differences to other coronaviruses, including the virus that caused the SARS epidemic in 2002–03. We also classified the proteins found in the virus into various families to help identify its functions. Discovering the functions of new proteins not only allows one to better understand their roles in their native contexts, but also utilize them in synthetic biology to assemble new biological circuits and pathways for useful applications such as production of valuable drugs for treating the disease.

In this study we explore the following research questions:

1. What are the characteristics of the genome sequence of COVID-19 coronavirus?

2. What are the similarities and differences between COVID-19 coronavirus, the bat-coronavirus, and SARS-coronavirus genome structures?

3. Is the COVID-19 coronavirus more similar to bat coronavirus or is it mutated SARS coronavirus?

4. Which class of the amino acid of the protein family accession does protein belong to?

For our analysis we use the Basic Local Alignment Search Tool (BLAST) algorithm to find similarity with other genome sequences. BLAST is one of the most heavily used sequence analysis tools available in the public domain [19, 20]. There is now a wide choice of BLAST algorithms that can be used to search many different sequence databases via the BLAST web page (http://www.ncbi.nlm.nih.gov/BLAST). The algorithm–database combinations can thus be executed either with default parameters or with customized settings. The results can also be viewed in various ways [21] and includes various calculations such as the highest score of the sequence, the total score across the genome, query coverage, expected value and percentage identity.

Since our aim is to find if the novel coronavirus has originated from bat or is mutated SARS coronavirus, we use deep learning methods to identify the protein accession families. Based on the outcome, we are able to understand the initial protein profile for the complete virus and our analysis could be useful to initiate drugs discovery research in newly identified organisms.

4. Experimental Set-up

4.1. Datasets

SARS-CoV-2 virus is a beta coronavirus, like MERS-CoV and SARS-CoV, both of which have their origins in bats. The Chinese rufous horseshoe bat *(Rhinolophus sinicus)* has been suggested to have direct lineage to the SARS coronavirus (SCoV), and the diversity of SARS-like CoVs (SLCoV) [22]. Hence the genomic and protein profiles of a COVID-19 and SARS causing coronavirus is compared to that isolated from a bat. The GenBank/EMBL/DDBJ accession numbers for the sequences obtained in our study are LR757996, DQ182595, and FJ588686. The genomic data used are coronavirus sequence assembly derived from patients affected in Wuhan in 2019-2020, patients infected with severe acute respiratory syndrome (SARS) in Zhejiang in 2003 and *Rhinolophus sinicus* species in 2006. The *SARS-CoV-2 strain (Genbank:* LR757996) is the coronavirus sequence assembly isolated from COVID-19 patient in 2019-2020 Wuhan outbreak. The SARS-COV strain (Genbank: DQ182595) is isolated from the throat swab from the first patient with SARS in Zhejiang. The bat SARS-CoV strain (Genbank: FJ588686) is isolated from *Rhinolophus sinicus* and is phylogenetically closer to human SARS-CoV than virus strains from other bat species

[18]. Viral sequences were downloaded from NCBI nucleotide sequence database [23].

Data Preparation

Protein profiling was conducted based on the pfam data of protein sequences and their family accession numbers. The dataset consists of 1.33 million protein sequences belonging to 17929 protein families collected from Kaggle. The data included sequences of length up to 2000 amino acids as depicted in the left sub-figure of Fig. 1.

We explore the sequence length of the protein sequences and plot the frequency of amino acids in Fig. 1. The histogram of the sequence length exhibits right skewness on count of protein sequences with the average sequence length of 155 amino acid and the median of 119 amino acid. For the purposes of this research, the sequences were limited to length of 119 amino acids to remove outliers. Also from the frequency of the amino acids in the dataset, we observe that the leucine (L) has the highest frequency in the whole dataset followed by Alanine (A) and Valine (V).

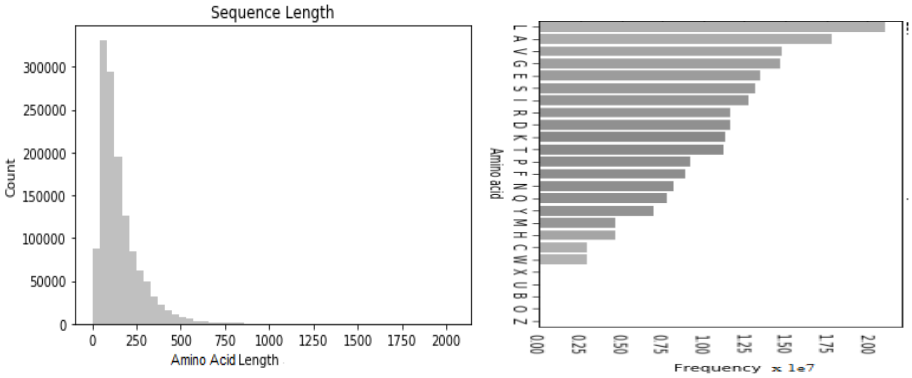


Fig. 1. Protein sequence length and frequency of Amino acids. Source: ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam32.0/Pfam-A.seed.gz, https://www.kaggle.com/googleai/pfam-seed-random-split

The dataset was split into training, validation and testing using the 80:10:10 ratio. For computational purposes, the number of families trained was limited to 1,000 common families which approximately correspond to 40% of the dataset. The protein sequences are integer coded for the exhaustive set of all the 20 amino acids [16].

4.2 Genomic Profiling

The Genomic Sequences are combinations of four nucleotides named: A(Adenine), T(Thymine), G(Guanine) and C(Cytosine). The combination of these sequences form the three base pair code for 20 different known amino acids. The amino acids sequences form the linear protein structure.

Subsequently, the genomes are compared for sequence similarity using the Basic Local Alignment Search Tool (BLAST). As the algorithm is based on finding local alignment, there may be multiple discrete regions of sequences with higher similarity score. BLAST is a sequence similarity search program that can be used via a web interface or as a stand-alone tool. There are several types of BLAST algorithms to compare all combinations of nucleotide or protein queries with nucleotide or protein databases. BLAST relies on heuristics to find short matches between two sequences and attempt to start alignments from these 'hot spots.' In addition to performing

alignments, BLAST provides statistical information to decipher the biological significance of the alignment; this is the 'expected' value, or false positive rate.

The Graphic Summary shows alignments of our query sequences. Fig. 2 represents how the similarity score is calculated. The aligned or positively matched sequence have a corresponding value and the gap represents a penalty. The difference between the summation of the values and summation of the penalties is used to compute the sequence similarity between genomes. The similarity score represents maximum sequence similarity.

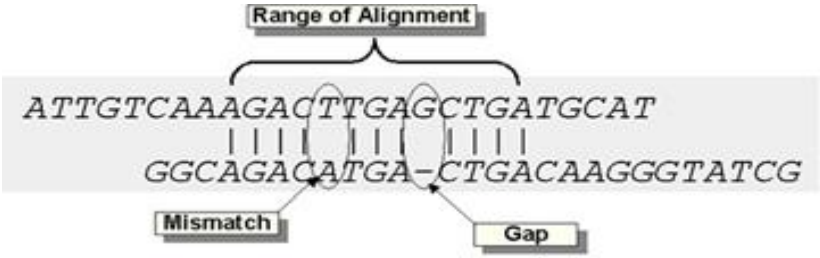


Fig. 2. Similarity Score

S is the sequence similarity score and is defined as follows:

$$S = \sum(identities, mismatches) - \sum(gap\ penalities) \quad (1)$$

$$Score = Max(S) \quad (2)$$

$$\%\ Coverage = \frac{pairwise\ alignment\ score\ of\ subject\ sequence(Score)}{length\ of\ query\ sequence} * 100 \quad (3)$$

4.3. Methodology

4.3.1. Gated Recurrent Unit (GRU)

Artificial Neural Networks (ANN) have been widely used in the field of text mining [24]. We have used Gated Recurrent Unit (GRU) [25] as the deep learning methodology in our study. GRU is a variation of Long Short-Term Memory (LSTM) and an improved variant of Recurrent Neural Networks (RNN) [26]. GRU has achieved remarkable performance in text classification problems [27]. Protein family is predicted from the unaligned protein domain sequence using a GRU due to its high success rate in identifying predominant protein family from the sequences [28, 29]. Our study can identify potential targets for combatting viral infection.

Neural Networks which is a subfield of artificial intelligence (AI) seek to build predictive models to classify or get insight from different types of data. Neural Network consists of different computational layers, including input, output layer, and hidden layers. The hidden layers are connected to input and output neuron on either side. An activation function is applied to generate output for an input. The architecture of Convolutional Neural Network (CNN) is different as it has a set of filters which scan the input for features irrespective of their position in the sequence. In Recurrent Neural Network, there is a time-delayed connection in between the neurons along with the feed forward neural network. For a biological sequence, RNN considers one sequence at a time and transfers information from output of previous step's hidden layer to input layer of next step. Moreover, in RNNs, the gradient of the loss function

exponentially decays over time (vanishing gradient). The GRU models are a varied type of RNN which is used for problems involving protein sequences [30, 31]. GRU has forget gates like LSTM architecture but with fewer parameters as GRU does not have output gate. The memory cells in LSTM allow it to learn longer-term dependencies. Along with it, it filters out the memory to store important input for many steps. There is, of course, a higher complexity and larger computer cost involved in the LSTM model. A GRU would be superior because it is simpler yet retains the ability to train for long-term dependency. GRUs are also easier to modify and can be trained faster than LSTMs. In this study we also present the Bi-LSTM result along with the GRU model performance.

In Fig. 3 we provided the internal architecture of GRU and Bi-LSTM models. In Figure 3(a) we show the building block of GRU where each building block can monitor the flow of information by updating and resetting gates revealing the memory at every step. The GRU attains the equilibrium between the previous and the next states of the memory as its output.

When both the input and output are sequences, we use bi directional GRU model which is an extended GRU model. In such cases, recurrence occurs in the input from both forward and backward directions of network. The use of bidirectional inputs can operate in two ways, one from past to future and one from future to past, allowing it to retain contextual knowledge from past and future at any time. In our study, the

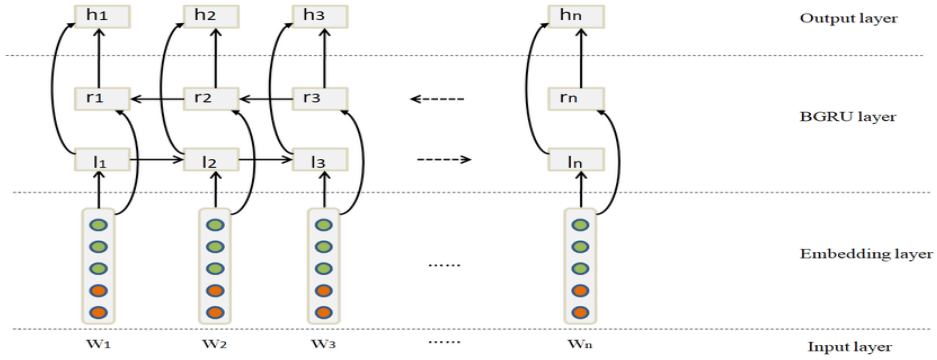


Fig. 3a. Architecture of GRU. Source: based on [32]

input is an amino acid sequence which can be processed from N-terminus to C-terminus and C-terminus to N-terminus. The hidden layer of this model handles the protein sequences by capturing the dependency information of subsequences from all the intermediate hidden values. The feature of the target subsequence in a cell is calculated based on its dependencies between the left neighboring subsequence and the right neighboring subsequence. This way the learning from the next step can be utilized to predict the earlier steps of the model. This leads to develop bidirectional RNN where neurons are divided into forward and backward layers. As sequencing and experimental characterization of data increase rapidly, bidirectional GRU could be useful over bi directional RNN, for discovery and prediction of similarity for a wide range of protein functions. Bidirectional GRU also fixes the problem of vanishing gradients inherent in regular RNNs.

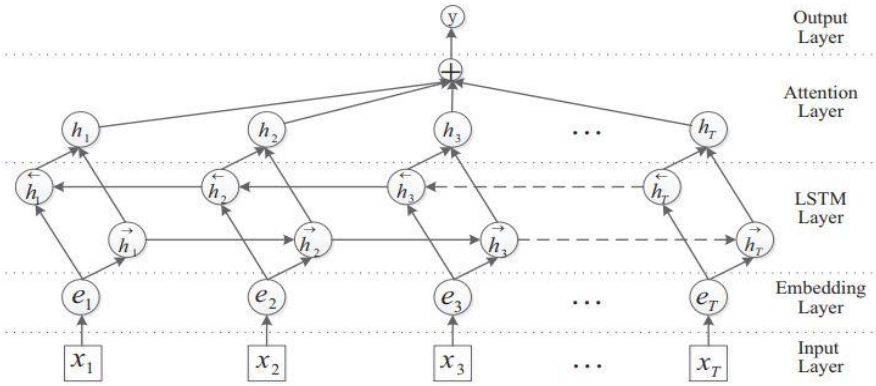


Fig. 3b. Architecture of Bi-LSTM Models
Source: https://github.com/kwonmha/Bidirectional-LSTM-with-attention-for-relation-classification

The mathematical abstraction of hidden layer activation in a neuron unit of GRU are defined as:

$$l_t = \sigma(W_l x_t + U_l h_{t-1} + b_l)$$
$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$
$$p_t = tanh(W_h x_t + U_h[r_t * h_{t-1}] + b_r)$$
$$h_t = (1 - l_t)h_{t-1} + lp_t$$

Where $l_t \rightarrow$ *update gate*
$r_t \rightarrow$ *reset gate*
$p_t \rightarrow$ *candidate activation vector*
$h_t \rightarrow$ *output memory state*
$h_{t-1} \rightarrow$ *output of Previous GRU block*
$W_x \rightarrow$ *weights of respective gate*(x) *neurons*
$\sigma \rightarrow$ *sigmoid function*
$U_x \rightarrow$ *recurrent weight matrics for gate*(x) *neurons*

The mathematical abstraction of hidden layer activation in a neuron unit of Bi-LSTM are defined as:

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o)$$
$$\hat{c}_t = tanh(W_c.[h_{t-1}], x_t) + b_c$$
$$c_t = f_t * c_{t-1} + i_t * \acute{c}_t$$
$$h_t = o_t * \tanh(c_t)$$

Where
i_t, f_t *and* o_t *denote input gate,* $forget\ gate$ and *output gate*
$\sigma \rightarrow$ *sigmoid function*
$W_x \rightarrow$ *weights of respective gate*(x) *neurons*
$h_{t-1} \rightarrow$ *output of previous LSTM block*$(at\ t-1)$
$x_t \rightarrow$ *input at current timestamp*
$b_x \rightarrow$ *biases for the respective gate*(x)
$c_t \rightarrow$ *cell state* $(memory)$ *at timestamp* (t)
$\acute{c}_t \rightarrow$ *candidate for cell state* $(at\ timestamp\ t)$

4.3.2. Model Implementation and Training

The GRU model is trained on 439,493 protein sequences using "Adam" optimizer for 10 epochs with the batch size of 256 sequences. The model is tested and validated with 54,378 sequences. The model architecture comprises of an

embedding layer, a dropout layer at 0.5 rate, a GRU layer with 1024 units and a last dense layer with 1000 neurons to classify into the 1000 families. As a multiclass classification problem, we classify a given amino acid sequence into one of 1000 protein families using the 'softmax' activation function. In order to compare their performances, we have used the same parameters to train the Bi-LSTM model.

5. Results and Discussion

5.1. Genome Similarity Analysis

The genome structure of the three viruses isolated from three different sources namely COVID patient, SARS patient, and the virus isolated from a bat are further partitioned into different base pairs that the DNA structure consists of. This base pairs define different pairing relationships that the DNA sequence generates. We can see from Table 2 constructed from exploratory analysis of the dataset, the COVID structure has a genome length of 29,868, followed by SARS structure with a length of 29,706, and finally by bat structure with a length of 29,059. Thus, all three structures have a substantial amount of genome length for our research to check for the similarity between their structures. Further, we studied the composition of the four nucleotides which make up the genome sequence. The genome wide exploration of the sequence showed little difference between the strains with close G-C content as well as the nucleotide percentages in the following order T>A>C>G. The similar G-C content is accounted for as a factor for the structural similarity between the strains.

Table 2. Table listing the genomic comparison

	Covid	SARS	Bat
Genome length	29,868	29,706	29,059
T(Thymine)	9,586	9,135	8,852
A(adenine)	8,933	8,450	8,266
G(Guanine)	5,861	6,184	6,085
C(Cytosine)	5,488	5,937	5,856
G-C Content	38.00%	40.80%	41.10%

We resorted to global alignment which is a sequence alignment over the entire length of two or more nucleic acid or protein sequences using Needleman-Wunsch algorithm [33]. Fig. 4 shows the pairwise alignment of the sequences based on global alignment algorithm. We found 82.42% identity match between the COVID-19 patient and the virus that has been isolated from a bat using the NCBI BLAST. Based on the NCBI similarity score of 27,375 using equation 1, we compute the coverage score of 91.6% between the genomes by equation 3.

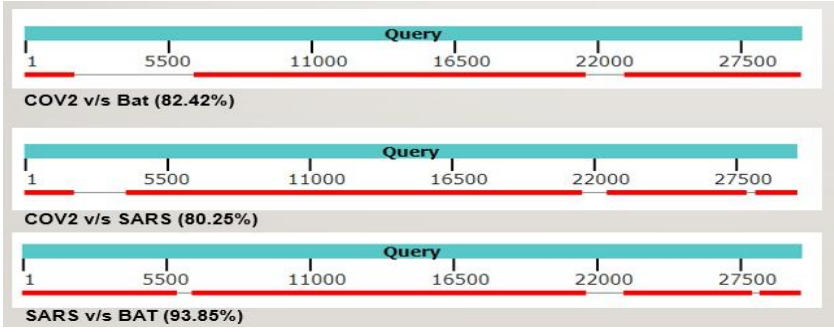


Fig. 4. Pairwise Alignment

In Fig. 4, all the highlighted red areas show different base pairs positions that have matched with most of the genome similarities between the viruses isolated from bats, SARS and the COVID-19 patient. In the top segment of figure 4, the highlighted red areas from 6000bp to 20000bp position on the genome show different base pairs positions that have matched with most of the genome similarities between the viruses that are isolated from bat and the COVID-19. In the middle segment of figure 4, the highlighted red areas from 4000bp to 20000bp show different base pairs positions that have matched with most of the genome similarities between the viruses isolated from SARS and the COVID-19 patients. From the above-received percentage similarities (82.42%), we can infer that CoV-2 has evolved from bat virus. Further the analysis of genome similarity between the viruses that are isolated from bats and the SARS patient as shown in the bottom segment of the figure 4, we observe a staggering similarity of 93.85%. We can see from the highlighted red areas that the entire genome structure is a match except for the areas between positions 6000bp to 6500bp and 21000bp to 23000bp.

5.2. Model Analysis

As depicted in Fig. 5, the proposed bidirectional GRU model is trained over the samples with known protein families from Pfam database and performed well through the epochs with no sudden spikes or drops in either training or validation. Further, there is a sharp significant increase and decrease in the accuracy and loss respectively at 2nd epoch. The accuracy curves, started from 18 % of training data and 91% of validation data in the first epoch. It goes up to 99% for training data and 98% for the validation data at 8th epoch (i.e. before reaching 10th epoch). On the other hand Figure 6 exhibits the curves for Bi-LSTM model, that started from 35% in the first epoch on training data and 69% on validation data, reached to a level of 94% and 96% for training and validation dataset respectively by the end of 10th epoch. Figure 5(b) shows that for GRU model there was significant levels of drop in loss values from 4.1 to 0.01 for training data and 0.34 to 0.06 for the validation data. Figure 6(b) shows the loss reduction for Bi-LSTM model from 3.3 to 0.247 for training data and from 1.6 to 0.129 for validation data respectively.

Fig. 5. Accuracy and Loss curves for GRU model

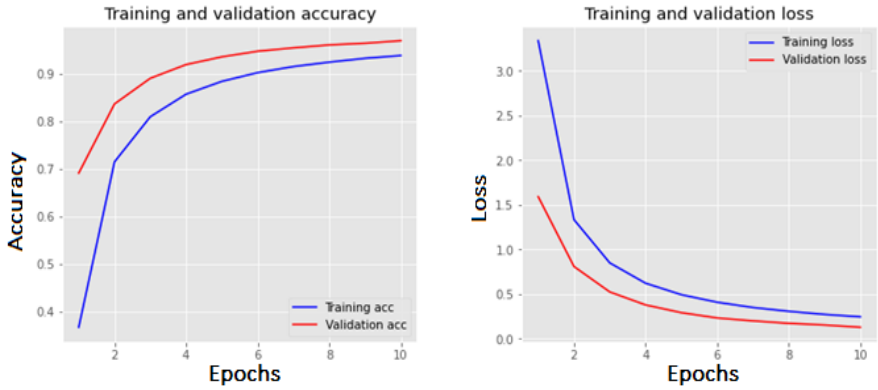


Fig. 6. Accuracy and Loss curves for Bi-LSTM model

Our proposed GRU model achieves training accuracy of 99.6%, validation and test accuracies of 98.2% compared to the Bi-LSTM model with training accuracy of 93.8%, validation and test accuracies of 96.6% as shown in Table 3. There is an improvement of 1.6% of our GRU model over the Bi-LSTM model for both the test and validation dataset and ≈6% improvement over training accuracy. The models use the Pfam database of all the 1000 protein families. This does not include the protein families belongs to SARS-CoV-2 as they are not identified for COVID-19 virus. Protein families are group of proteins that share a common evolutionary origin, reflected by their related functions and similarities in sequence or structure. Thus, the acquired GRU model is capable of finding different protein family accessions present in different genome sequences that are included in Pfam database. Since the GRU model uses the unaligned protein sequences to predict families, by predicting the protein families of CoV proteins, we will be able to identify functionally similar proteins. These functionally similar proteins can give an insights on the pathogenicity of the viruses.

Table 3. GRU and Bi-LSTM Model Performance Metrics

	TRAIN		**TEST**		**VALIDATION**	
Model	GRU	Bi-LSTM	GRU	Bi-LSTM	GRU	Bi-LSTM
Data Size	439493	439493	54378	54378	54378	54378
Accuracy	99.6%	93.8%	98.2%	96.6%	98.2%	96.6%
Loss	0.01	024	0.06	0.13	0.06	0.13

The CoV-2 protein sequences are translated from the DNA sequences downloaded from NCBI. These CoV-2 protein sequences generated are used to predict the protein families. We got an average of 76% probability as the measure of confidence for the prediction of protein families included in CoV-2. We computed the result from the average probabilities from the probability distribution against the protein families generated by the Softmax function used in the trained model. Our model classifies protein family accessions which are respectively present in the genome structure of SARS-COV-2, SARS-COV and bats. Table 4 defines the count of total amino acids, total proteins, number of functional proteins, and the number of family classes present in each classified genome structure.

5.3. Protein Family Accession Analysis

Table 4 exhibits the profile of the protein sequences of three different strains. Further stacking of protein family accession based on the total number of occurrences

gives a clear distinction of which protein family accession is a prominent member of all three viruses as shown in Figure 7. We found that the PF00560.3 family is the prominent protein family accession in all three viruses. PF00560.3 is a family of leucine-rich repeats and a similar motif has been discovered to play an important role in pathogenesis of SARS-CoV [34]. The presence of similar proteins indicates nucleocapsid protein's role as a shuttle protein responsible for transporting viral genome across membranes [35]. This method provides us with a basis to know which protein family is a prominent member. which can be targeted for rendering a virus incapable of affecting a human host.

Table 4. Protein profiles of the 3 strains

	SARS-COV-2	**SARS**	**Bat**
Genome length	29868	29706	29059
Total amino acids	9956	9902	9686
Total proteins	747	728	699
Functional proteins	89	84	76
Number of family classes	41	44	42

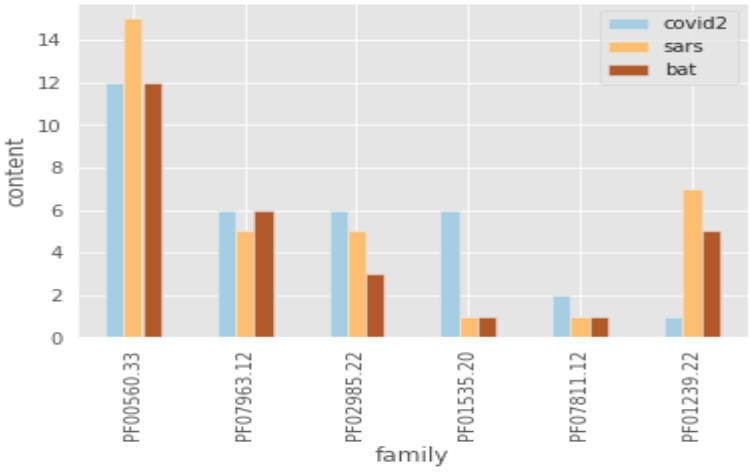


Fig. 7. Pfam content in each genome

Though Bi-LSTM is superior to RNNs in this application as Bi-LSTM could retain the "memory" of outcomes of the previous layers, but they are inferior to GRU due to its higher complexity and computer cost. GRUs are simpler yet retain retains the ability to train for long-term dependency, easier to modify and can be trained faster than LSTMs [26, 36] with same number of trainable parameters. Moreover GRUs show better performance for smaller and less frequent dataset [37, 38]. In our experiment we show that GRU outperforms Bi-LSTM in terms of accuracy of multi-class protein classification and the model loss criterion. Exploration of Temporal Convolutional Network (TCN) is another avenue to successfully model the protein sequences [39] where TCN used significantly lesser memory than LSTM and GRUs to store partial results. Bai et al. [40] showed that TCN outperformed recurrent architectures in modeling gene sequence for large genetic dataset. In our research we focus on GRU as the methodology for classification and compared the performance with that of Bi-LSTM model. We will explore TCN in our application as future works.

6. Conclusion

The methodology and the proposed classification model discussed above are based on the complete genome data of an organism. The initial genome sequence was released by the Chinese authorities in mid-January, 2020 as detected in the first patients. This followed by other institutions releasing the genomic sequences starting late January 2020. The outcomes of our study demonstrate that the CoV-2 virus genome structure is more similar to the bat virus than the SARS virus. The results show that there is 82.42 % match between CoV-2 and bat virus genome structures. In addition, we have found that PF00560.3 is the protein family accession that is prominent in all the virus structures. Our GRU model predicts PF00560.3 as the prominent protein family for CoV-2 with an average probability of 76% computed from the probability distribution generated by the Softmax function output against the protein families. Our finding infers that CoV-2 proteins turns out to be functionally similar to the proteins found in previous strains as observed in prior research by Reed et al. [25].

In this study we self-imposed few limitations to restrict the GRU model to classify protein families with only top 1000 protein families due to considerable computational resources required to train and test the model. Nevertheless, the results are notable and provides insights to understand coronavirus and its protein classification. The results can further be employed in medical research, considering the functionalities of PF00560.3 and how they are responsible for affecting a human host. In addition, our results identify the protein family that can be used to develop counterfeit measures against the Coronavirus.

In future work we would suggest to carry out a comparative study with the annotated sequences to power up our existing model. We would work to extend our classifier and improve its accuracy with optimized hyperparameters to include all protein families. We also intend to study the performance of our model using one versus rest (binary) classifier for the prediction of the protein family in our future work. Additional models like TCN with hyperparameter tuning will be explored in future in addition to GRU and Bi-LSTM model presented in this research.

Acknowledgments

We acknowledge Ramya Chimata Venkatakrishnan and Venkat Nihaal Akula for their help in this project. Any remaining errors are authors' responsibility.

References

1. Lu R., Zhao X., Li J., Niu P., Yang B., Wu H., et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 395(10224), 565-574 (2020). DOI: https://doi.org/10.1016/S0140-6736(20)30251-8.
2. Guo Y.-R., Cao Q.-D., Hong Z.-S., Tan Y.-Y., Chen S.-D., Jin H.-J., et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. Military Medical Research, 7(1), 1-10 (2020). DOI: https://doi.org/10.1186/s40779-020-00240-0.
3. Ruan Y., Wei C. L., Ling A. E., Vega V. B., Thoreau H., Thoe S. Y. S., et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. The Lancet, 361(9371), 1779-1785 (2003). DOI: https://doi.org/10.1016/S0140-6736(03)13414-9.
4. Fehr A. R., Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Coronaviruses. Methods of Molecular Biology, 1282, 1-23 (2015). DOI: 10.1007/978-1-4939-2438-7_1.
5. Wu F., Zhao S., Yu B., Chen Y.-M., Wang W., Song Z.-G., et al. A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269 (2020).

6. Zhou P., Yang X.-L., Wang X.-G., Hu B., Zhang L., Zhang W., et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), 270-273 (2020).
7. De Wit E., Van Doremalen N., Falzarano D., Munster V. SARS and MERS: recent insights into emerging coronaviruses. Nature Reviews Microbiology, 14(8), 523-534 (2016).
8. Wu A., Peng Y., Huang B., Ding X., Wang X., Niu P., et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell host & Microbe, 27(13), 325-328 (2020). DOI: https://doi.org/10.1016/j.chom.2020.02.001.
9. Angeletti S., Benvenuto D., Bianchi M., Giovanetti M., Pascarella S., Ciccozzi M. COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. Journal of medical virology, 92(6), 584-588 (2020). DOI: https://doi.org/10.1002/jmv.25719.
10. Tang X., Wu C., Li X., Song Y., Yao X., Wu X., et al. On the origin and continuing evolution of SARS-CoV-2. National Science Review, 7(6), 1012–1023 (2020). https://doi.org/10.1093/nsr/nwaa036.
11. Zhu N., Zhang D., Wang W., Li X., Yang B., Song J., et al. A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine, 382(8), 727-733 (2020). DOI: 10.1056/NEJMoa2001017.
12. Bileschi M. L., Belanger D., Bryant D. H., Sanderson T., Carter B., Sculley D., et al. Using deep learning to annotate the protein universe. bioRxiv, 1-28. (2019). DOI: https://doi.org/10.1101/626507.
13. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, 521(7553), 436-444 (2015).
14. Bateman A., Coin L., Durbin R., Finn R. D., Hollich V., Griffiths-Jones S., et al. The Pfam protein families database. Nucleic acids research, 32(suppl_1), D138-D141 (2004). DOI: https://doi.org/10.1093/nar/gkh121.
15. D'Agaro E. Artificial intelligence used in genome analysis studies. The EuroBiotech Journal, 2(2), 78-88 (2018). DOI: https://doi.org/10.2478/ebtj-2018-0012.
16. Vijay R. Protein Sequence Classification: A case study on Pfam dataset to classify protein families. https://towardsdatascience.com/protein-sequence-classification-99c80d0ad2df. Last accessed 2019/09/02.
17. Hu H., Li Z., Elofsson A., Xie S. A Bi-LSTM based ensemble algorithm for prediction of protein secondary structure. Applied Sciences, 9(17), 3538 (2019).
18. Jurtz V. I., Johansen A. R., Nielsen M., Almagro Armenteros J. J., Nielsen H., Sønderby C. K., et al. An introduction to deep learning on biological sequence data: examples and solutions. Bioinformatics, 33(22), 3685-3690 (2017). DOI: https://doi-org.ezaccess.libraries.psu.edu/10.1093/bioinformatics/btx531.
19. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410 (1990). DOI: https://doi.org/ 10.1016/S0022-2836(05)80360-2.
20. Ye J., McGinnis S., Madden T. L. BLAST: improvements for better sequence analysis. Nucleic acids research, 34(suppl_2), W6-W9 (2006). DOI: https://doi.org/10.1093/nar/gkl164.
21. McGinnis S., Madden T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research, 32(suppl_2), W20-W25 (2004). DOI: https://doi.org/10.1093/nar/gkh435.
22. Yuan J., Hon C.-C., Li Y., Wang D., Xu G., Zhang H., et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. Journal of general virology, 91(4), 1058-1062 (2010). DOI: https://doi.org/10.1099/vir.0.016378-0.
23. Wheeler D. L., Barrett T., Benson D. A., Bryant S. H., Canese K., Chetvernin V., et al. Database resources of the national center for biotechnology information. Nucleic acids research, 36(suppl_1), D13-D21 (2007). DOI: https://doi.org/10.1093/nar/gkm1000
24. Song S., Huang H., Ruan T. Abstractive text summarization using LSTM-CNN based deep learning. Multimedia Tools Applications, 78(1), 857-875 (2019). DOI: https://doi.org/10.1007/s11042-018-5749-3
25. Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., et al. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. in EMNLP, Association for Computational Linguistics, pp. 1724-1734.
26. Sherstinsky A. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. Physica D: Nonlinear Phenomena, 404(132306), 1-28 (2020). DOI: https://doi.org/10.1016/j.physd.2019.132306.
27. Zulqarnain M., Ghazali R., Ghouse M. G., Mushtaq M. F. Efficient processing of GRU based on word embedding for text classification. International Journal on Informatics Visualization, 3(4), 377-383 (2019). DOI: 10.30630/joiv.3.4.289

28. Lee T. K., Nguyen T. Protein family classification with neural networks. Stanford University, pp. 1-9 (2016).
29. Le N. Q. K., Yapp E. K. Y., Nagasundaram N., Chua M. C. H., Yeh H.-Y. J. C. Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture. Computational and Structural Biotechnology, 17, 1245-1254 (2019). DOI: https://doi.org/10.1016/j.csbj.2019.09.005.
30. Pfeiffenberger E., Bates P. A. Predicting improved protein conformations with a temporal deep recurrent neural network. PLos One, 13(9), e0202652 (2018). DOI: https://doi.org/10.1371/journal.pone.0202652.
31. Le N. Q. K., Yapp E. K. Y., Yeh H.-Y. ET-GRU: using multi-layer gated recurrent units to identify electron transport proteins. BMC Bioinformatics, 20(1), 1-12 (2019).
32. Zhao M., Wang H., Guo J., Liu D., Xie C., Liu Q., et al. Construction of an industrial knowledge graph for unstructured chinese text learning. Applied Sciences, 9(13), 2720. (2019). DOI: 10.3390/app9132720.
33. Needleman S. B., Wunsch C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol, 48, 443-153 (1970). DOI: https://doi.org/10.1016/0022-2836(70)90057-4.
34. Reed M. L., Howell G., Harrison S. M., Spencer K.-A., Hiscox J. A. Characterization of the nuclear export signal in the coronavirus infectious bronchitis virus nucleocapsid protein. Journal of virology, 81(8), 4298-4304 (2007). DOI: https://doi.org/10.1128/JVI.02239-06.
35. Timani K. A., Liao Q., Ye L., Zeng Y., Liu J., Zheng Y., et al. Nuclear/nucleolar localization properties of C-terminal nucleocapsid protein of SARS coronavirus. Virus research, 114(1-2), 23-34 (2005). DOI: https://doi.org/10.1016/j.virusres.2005.05.007
36. Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10) 2451-2471 (2000). DOI: https://doi.org/10.1162/089976600300015015.
37. Chung J., Gulcehre C., Cho K., Bengio Y. (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. NIPS Workshop on Deep Learning. DOI: http://arxiv.org/abs/1412.3555.
38. Gruber N., Jockisch A. Are GRU cells more specific and LSTM cells more sensitive in motive classification of text? Frontiers in Artificial Intelligence, 3, 1-6 (2020). DOI: https://doi.org/10.3389/frai.2020.00040.
39. Kim H. Y., Kim D. Prediction of mutation effects using a deep temporal convolutional network. Bioinformatics, 36(7), 2047-2052 (2020). DOI: https://doi.org/10.1093/bioinformatics/btz873.

Comparing Pregnancy and Childbirth-related Hospital Visits in Arizona Before and During COVID-19 Using Network Analysis

Jinhang Jiang[1[0000-0002-3941-5161]], Karthik Srinivasan[2[0000-0001-6094-4722]]

[1] Walmart International Technology, Bentonville AR, U.S.;
[2] School of Business, University of Kansas, Lawrence KS, U.S.

https://doi.org/10.33847/2686-8296.3.2_2

Received 01.10.2021/Revised 09.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. The COVID-19 pandemic has had a severe effect on all facets of human society, including healthcare. One of the primary concerns in healthcare is understanding and mitigating the impact of the pandemic on pregnancy and childbirth. While several studies have looked at challenges such as contract tracing of positive cases, predicting confirmed cases and deaths in individuals and communities, few studies have examined differences in hospitalization and treatment of pregnant mothers and infant care in large populations. In this study, the prevalence and co-occurrence of pregnancy and childbirth-related diagnoses reported in Arizona State hospitals for three sixth-month periods - before COVID-19 (second half of 2019), COVID-19 onset (first half of 2020), and COVID-19 (second half of 2020) are analyzed using network analysis. The results show that there are considerable differences in ego networks of few diagnoses during these time periods warranting further investigation into the causality of such population changes.

Keywords: Health analytics, Network analysis, Ego networks, Exploratory data analysis (EDA), COVID-19 data analysis, Pre-term and post-term conditions, Pregnancy and childbirth, Newborn diagnoses**.**

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, also known as coronavirus or COVID-19, was first identified in Wuhan, China, in December 2019 and has spread to almost all the countries since. It was determined as a global pandemic by WHO on March 11th, 2020. As of June 1st, 2021, more than 170 million cases have been confirmed, with more than 3.73 million confirmed deaths attributed to COVID-19, making it one of the deadliest pandemics in history.

COVID-19 related research is being conducted in multiple disciplines, including physiology, pharmacology, epidemiology, economics, political sciences, computer science, and so on. Using the vast amount of data being generated in information systems, the field of data science or analytics has become invaluable in testing scientific theories, generating new knowledge, and solving complex problems. Data science or analytics is the systematic computations analysis for discovering, interpreting, and communicating meaningful patterns in data. Multiple analytics studies have been conducted to discern patterns, determine risks, and make a reliable prediction using COVID-19 data [1–3]. However, only a few studies have been done on analyzing the impact of the pandemic on pregnancy and newborn health and wellbeing [4]. In this study, the focus is on exploring the inpatient visits discharge data from public hospitals of Arizona, USA, for differences in diagnoses reported for pregnancy and newborn health conditions before and during the COVID-19 pandemic.

Real-time dashboards and reports such as those published by the Arizona

department of health services (AZDHS) on their website[1] typically run database queries and provide summaries either in tables or chart visualizations (e.g., bar chart, heatmaps, etc.). While these are helpful to understand the overall patterns in COVID-19 confirmed cases and deaths, their scope in knowledge discovery of linked entities and uncovering phenomena such as disease co-occurrence is limited. This study reports distinct differences in the prevalence (i.e., the total number of times a diagnosis is reported in a sample population) of pregnancy and newborn diagnoses across three sixth-month periods - before COVID-19 (second half of 2019), COVID-19 onset (first half of 2020), and COVID-19 (second half of 2020). Thereafter, networks capturing the co-occurrence of diseases (i.e., disease co-occurrence network) are created for each of these three time periods to enable novel knowledge discovery. Subsets of these networks focused on individual disease nodes called ego networks are presented to show how diagnoses are reported differently before and during the pandemic. The analysis shows that there are specific disease diagnoses that have pronounced differences in their prevalence as well as co-occurrence with other diseases, thus warranting further investigation. Our study has multiple implications for policy making and provides directions for setting hypotheses for causal analysis and clinical discovery. It provides a novel approach to exploring the list of diagnoses reported in hospital visits discharge data or large populations.

2. Literature Review

The COVID-19 pandemic became global in early 2020, but it is nowhere close to the end. Studies have focused on modeling of COVID-19 spread [1–3], risks [3], and early prediction [2]. Person-level factors, including age, gender, existing health conditions, awareness, and socio-economic status, have been associated with the individual risk of contracting COVID-19 [5,6]. Hypothesis testing, regression modeling, standard data mining methods are helpful to answering specific research questions and making predictions. However, exploratory data analysis (EDA) of complex datasets often requires novel methodologies specific to the problem and application. Specifically, patient visits and discharge data collected across multiple hospitals contain a multitude of information such as physician notes, patient information, insurance, payment details, diagnoses codes, procedures, lab reports, and medications. There is a need to utilize novel analytics methodologies such as network analysis to examine linked entities such as diagnoses codes as they are reported simultaneously in visit summaries using standard reporting code.

Network analysis is a research approach used for analyzing the structural and relational aspects of linked data represented as graphs. There are several methods in network analysis such as network data visualization [7], descriptive analysis of the structural properties of the graph and individual properties (e.g., page rank, eigenvector centrality, clustering coefficient, density, etc.) [8], network diffusion, subgraph analysis and community detection [9], mathematical modeling of the generative process of the network structure (e.g., preferential attachment mechanism, Erdos-Renyi model, etc.) [10], statistical modeling (ERGM modeling) [11], and knowledge graph embedding [12]. Few studies have looked at the network of diseases reported in hospital patient visit records [13–15]. This study adopts the disease co-occurrence network proposed by [13] to compare occurrence and co-occurrence patterns of pregnancy and newborn-related diagnoses in Arizona hospital inpatient visit records before and during the COVID-19 pandemic. While a few studies have looked at the COVID-19 pandemic affecting pregnancy and newborn [4,16,17], there is still a lot to understand about the immediate and long-term effects of the pandemic on pregnancy and newborn health and wellbeing. Therefore, an exploratory data analysis (EDA)

[1] https://www.azdhs.gov/covid19/

using a novel network analysis approach used in this study can help unearth previously unknown patterns and provide a foundation for future research in causal exploration and knowledge discovery in this field.

According to the U.S. Center for Disease Control (CDC), COVID-19 is uncommon in newborns born to mothers who had COVID-19 during pregnancy [18]. Most newborns who tested positive for COVID-19 had mild or no symptoms, and they easily recovered. However, even though infants are less likely to be infected by COVID-19 than adults, it is not established if they are really at low risk. COVID-19 might not affect infants directly, but the pandemic could have indirect impacts on the hospitalization and the care for pregnant and newborn patients. For example, limited beds and resources during the peak of the pandemic had an impact on non-COVID-19 hospital admissions, including those related to child delivery. The public hospital discharge records are an ideal resource for studying individual and population-level impacts of the pandemic on pregnancy and newborn health and wellbeing. Though each patient cannot be individually screened and analyzed, systematic patterns in diagnoses record inpatient visits across public hospitals in a large state such as Arizona can unravel patterns about the direct and indirect impact of COVID-19 pandemic on pregnancy and newborn hospital visits and quality of care.

3. Data and Methodology

3.1. Data

In this study, the Arizona Hospital Discharge Data Public Use Files (PUF) provided by the Arizona Department of Health Services (AZDHS)[2] was used as the data source to conduct the analysis. Since the data is public use and governed by U.S. federal privacy laws for published health data, no personal identification data is provided, and each visit record is independent from other records. Other precautions such as removal of time of visit and location of the hospital are also carried out in the PUF datasets. The health conditions and diseases for treatment are recorded using the International Classification of Diseases (ICD-10) nomenclature, a well adopted standardized nomenclature across healthcare providers in the world for disease reporting in digital (electronic) health records. For this study, only the ICD-10 diagnoses codes of patient visit discharge records are used for disease network creation EDA. Similar Electronic Health Records (EHR) as the one used in this analysis can be publicly accessed, such as the MIMIC and eICU datasets [19], from state and national health agencies, aggregated data curators and vendor websites such as HCUP[3]. Though ICD-10 has been used in this study, the prevalence analysis and disease co-occurrence network analysis can be carried out using other standard codes such as ICD-11 or SNOMED diagnoses codes. More information about ICD-10 coding and fields in the PUF data used in this study can be found in the respective public information websites[4,5].

Particularly, the inpatient visits records during three six-month periods - before COVID-19 (second half of 2019), COVID-19 onset (first half of 2020), and COVID-19 (second half of 2020) were independently examined and compared. We refer to the three datasets as 2019-02 or pre-pandemic, 2020-01 or early pandemic, and 2020-02 or peak outbreak period in the rest of the paper. 2019-02, 2020-01, 2020-02

[2] https://azdhs.gov/preparedness/public-health-statistics/hospital-discharge-data/index.php#data-release

[3] https://www.hcup-us.ahrq.gov/

[4] https://www.icd10data.com/

[5] https://azdhs.gov/documents/preparedness/public-health-statistics/hospital-discharge-data/data-release/PUF_data_dictionary.pdf

datasets contain 382,334, 359,902, and 371,315 patient visits, respectively. A representative snapshot of how the data looks like is provided in Figure 1. Each patient visit record (i.e. row) can vary between 1 to 25 diagnoses codes. For example, in Figure 1, the admitting diagnoses is M79605 – Pain in the left leg, and the two subsequently reported diagnoses are N186 – End-stage renal disease and D631 – Unspecified lump in the right breast.

ADMITTING_DIAGNOSIS	AGEYG	DIAGNOSIS_2	DIAGNOSIS_3	DIAGNOSIS_4	DIAGNOSIS_5	DIAGNOSIS_6
M79605	70-79	N186	D631	N2581	Z6842	I120
N179	80-89	G909	G309	G20	E785	E860
R1013	30-39	K766	K7031	F1011	I10	R739
R55	70-79	N179	R55	E039	G8929	I10
R0600	60-69	I2510	I10	E785	F329	F419
S065X9A	90+	G935	Z515	Z66	S12000A	S12100A
I639	70-79	E1122	I129	N183	E785	H02401
E872	40-49	A419	G92	E872	R45851	F10239
R4182	60-69	J9621	G9341	I21A1	I63411	J189
R918	70-79	F10239	E871	I10	D630	D72829
R4182	70-79	G9341	Z66	D631	E871	N179
R531	80-89	J9611	J90	K5900	R1030	R531
F05	70-79	U071	F0150	F0280	F419	F329
R0602	60-69	N170	I5031	I161	E871	N189
G92	70-79	R4701	G40209	Z7982	T421X5A	R51
K5732	40-49	I10	E282	E119	Z90710	K219

Fig. 1. Snapshot of ICD-10 diagnoses reported in the hospital visit discharge records

3.2. Creating a Disease Network

Similar to genome-genome networks or protein-protein networks [20], disease (co-occurrence) networks can be created based on the evidence that pairs of diseases were simultaneously reported in multiple patient visits. Such a network is typically an undirected weighted graph represented as $G(V,E,W)$ where $G(.)$ is the notation for the graph, V is the set of diagnoses codes representing diseases, E is the set of links or edges between pair of diseases $v_1, v_2 \in V$ representing co-occurrence, and W is the set of weights assigned to the edges, typically a function of the number of times the pairs of diseases co-occurred across patient visits. In this study, the disease co-occurrence network proposed by [13] with edge weight defined as follows:

$$w_{xy} = \frac{\sqrt{2}C_{xy}}{\sqrt{{P_x}^2 + {P_y}^2}}$$

Where, C_{xy} is the co-occurrence of diseases x and y across patient encounters, P_x and P_y are prevalence of diseases x and y respectively. w_{xy} is symmetric, reflexive, and constrained to the range [0,1] and therefore is more tractable as an edge weight than C_{xy} or other proposed edge weights [14].

3.3. Pre-processing

As part of pre-processing, diagnoses codes occurring less than ten times in 6 months (i.e., rare diseases) were screened out. In total, 8972 ICD-10 were considered for the disease network creation. A total of 4,353,903 edges were in the disease network. Thus, the average weighted degree is 970.55. The high average weighted degree is attributed to common diseases such as hypertension and diabetes that are not only highly prevalent but also commonly co-occur with other conditions.

For this study, the focus is on diagnoses codes listed in Table 1 for generating ego-networks. Ego network for any disease contain the disease node and all the diseases directly linked to it along with the respective edges with a weight bar of 0.05.

Table 1. ICD-10 codes related to pregnancy and newborn that are of focus in this study

ICD-10 Code	Description	Prev_2019 2	Prev_2020 1	Prev_2020 2
z370	Single live birth	38772	35075	37060
z371	Single stillbirth	294	292	302
z381	Single liveborn infant, born outside hospital	134	117	120
z20828	Contact with and (suspected) exposure to other viral communicable diseases	69	17474	96083
o76	Abnormality in fetal heart rate and rhythm complicating labor and delivery	5365	4821	5739
o80	Encounter for full-term uncomplicated delivery	5954	4701	3790

These six starting nodes allowed us to study how the disease networks for live birth, stillbirth, infant born outside a hospital, complicating labor and delivery, and uncomplicated (normal) delivery changed during COVID-19. The ego-networks of the first two ICD-10 codes should give us an idea of how the disease networks changed for the baby who was born in the hospital. The ego network for ICD-10 code "z381" should give us insight into how the disease network changed for the baby born outside the hospital. The diagnosis "o76" has the highest prevalence across the abnormal, complicated delivery, which is supposed to help us understand the negative impact the pandemic environment made on the process of delivering. The investigation of code "o80" can help understand how COVID-19 affects pregnant mothers' maternal health conditions as well as the uncomplicated delivery. "z20828" is the ICD-10 code directly related to COVID-19. Even though ICD10 "u071" is the code specifically created for COVID-19 during early 2020, since it was not available earlier, most patients having COVID-19 were labeled with "z20828" instead of u071 upon hospital admission. Therefore, "z20828" is used as an indicator of a (suspected) COVID-19 related admission.

A lower threshold of 0.05 is set for the edge weights in each network to reduce the density and find the most relevant results. This floor value reduces the total number of edges by more than 99.5 percent. For the analysis, Google Colab is used as the interactive Python code editor and the NetworkX and Matplotlib packages are used for generating and analyzing network objects and creating the data visualizations. While investigating the graphs, abnormal patterns of the networks were extracted to subgraphs to have a closer look. This method is helpful to track patterns of changes in the disease network over the three periods.

4. Results and Inferences

Prevalence/Frequency Analysis

Figure 2 shows the prevalence of each disease of focus in three different stages of the pandemic. The second from top-left is the ICD-10 code "z20828" related to early COVID-19 symptoms, and the patterns of the pregnancy and newborn diagnoses can be compared with this plot. The prevalence of all five targeted diseases dropped from the pre-pandemic stage to the early-pandemic stage. It is a signal that in the early-pandemic stage, the hospitals were not ready to handle the increasing cases of COVID-19 related diseases, resulting in the fact that they underreported the "other health services," such as pregnancy. The number of stillbirths and abnormality in fetal heart rates is still increasing in the second half of 2020, indicating the worst is yet to come. Moreover, the number of uncomplicated deliveries keep decreasing despite the

reported increase in the number of single live birth deliveries from the first half to the second half of 2020.

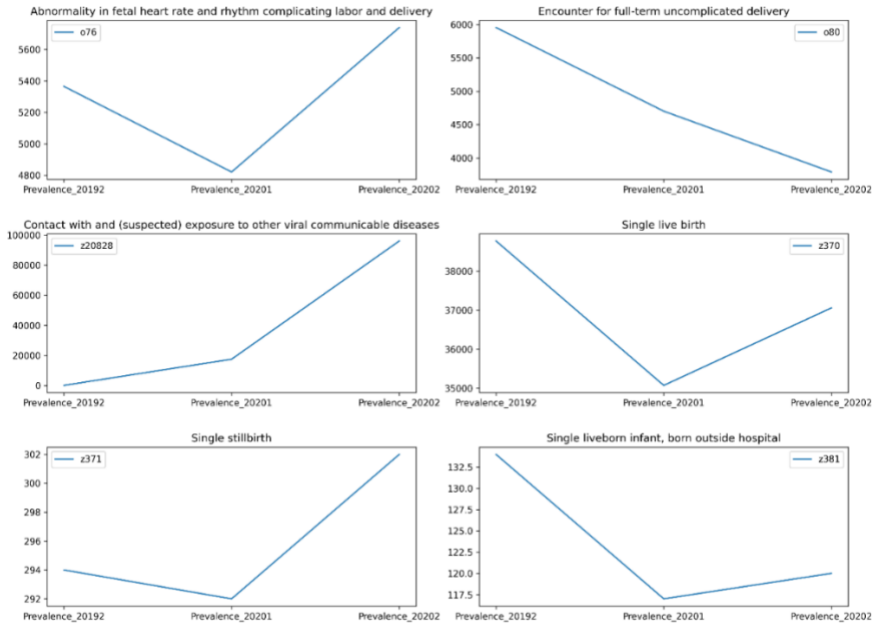
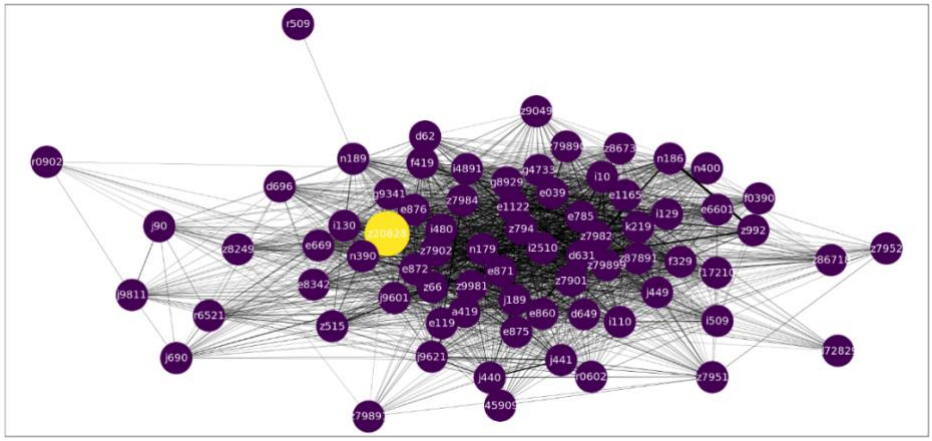
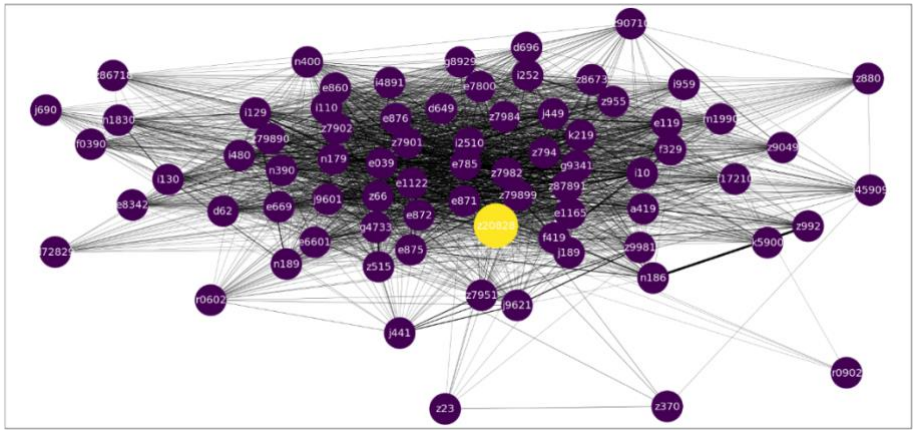


Fig. 2. Change in prevalence of diseases/conditions across three time periods Ego-network Analysis

In Figure 3, (a) and (b) have 64 common nodes. The unique nodes in each ego networks are shown in the "Unique diagnoses" column in Table2. As this study focuses on the impact of Covid19 on hospital visits with pregnancy and newborn related diagnoses, the "z370", ICD-10 code for single live birth, the first time presented in the ego network of "z20828" and underlined that the Covid-19 indeed had direct effects on the pregnancy and newborn.



(a)



(b)

Fig. 3. Ego networks of z20828 - Contact with and suspected exposure to other viral communicable diseases for (a) 2020-01, and (b) 2020-02

Fig. 3 (a), the ego network of "z20828" during the period of Covid-19 onset (first half of 2020), has 74 nodes (distinct diagnoses) and 1899 edges. According to the column of "Category" in Table 2, the 74 diagnoses can be split into 11 categories based on the starting alphabet of the ICD-10 codes. Moreover, among the 11 categories, the top 5 - most popular diagnosis categories are the "Other health services", which accounted for 20 nodes in the ego network. It is followed by "Endocrine, nutritional and metabolic diseases", "Diseases of the respiratory system", "Diseases of the circulatory system", and "Blood and immune disorders", which accounted for 13 nodes, 10 nodes, 8 nodes, and 5 nodes, respectively in the ego network. Furthermore, the average weight is 0.0992 for the ego network.

Fig. 3 (b) is the ego network of "z20828" during the second half of 2020, which has 75 nodes and 1993 edges. While the number of the nodes did not change much, the number of edges of Figure 3 (b) increased roughly 5 percent from Fig. 3 (a), a signal that during the second half of 2020, people with Covid-19 were more likely to have other diagnoses simultaneously. The 75 diagnoses, as shown in Table 2, can be split into 12 categories based on the starting alphabet of the ICD-10 codes. In spite of that, Fig. 3 (b) maintained the 11 categories from Figure 3 (a), the new category – m - Diseases of the musculoskeletal system and connective tissue joined the list of diagnosis categories. Moreover, the top 5 - most popular diagnosis categories in Fig. 3 (b) are "Other health services," "Endocrine, nutritional and metabolic diseases," "Diseases of the circulatory system," "Diseases of the respiratory system," and "Diseases of the genitourinary system", which accounted for 22 nodes, 14 nodes, 9 nodes, 7 nodes, and 6 nodes, respectively. The average weight for edges in Figure 3 (b) is 0.0994 which is a slightly increase from Fig. 3 (a).

Table 2, made for the purpose of comparison, displays the information of the diagnoses in the ego networks in Fig. 3. Table 2 describes the prominent ICD-10 diagnoses identified in the ego network of "z20828." The *m* category (i.e., diseases of the musculoskeletal system and connective tissue) of ICD-10 appear in the network corresponding to the second half of 2020. A possible reason for this peculiar pattern could be that hospitals may have filled with capacity with COVID-19 related cases and other critical admitting diagnoses, while postponing visits to chronic ailments such as osteoarthritis as much as possible.

Table 2. Information of the ego networks of "z20828"

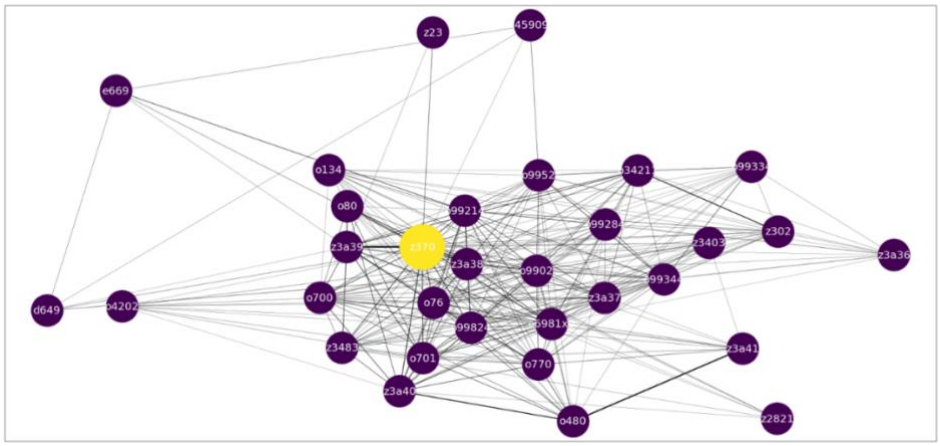
Time period	Category	Unique diagnosis
Covid-19 onset (first half of 2020)	a - Certain infectious and parasitic diseases d - Blood and immune disorders e - Endocrine, nutritional and metabolic diseases f - Mental and behavioral disorders g - Diseases of the nervous system i - Diseases of the circulatory system j - Diseases of the respiratory system k - Diseases of the digestive system n - Diseases of the genitourinary system r - abnormal clinical findings z - Other health services	*d631* - Anemia in chronic kidney disease *i509* - Heart failure, unspecified *j440* - Chronic obstructive pulmonary disease with (acute) lower respiratory infection *j90* - Pleural effusion, not elsewhere classified *j9811* – Atelectasis *r509* - Fever, unspecified *r6521* - Severe sepsis with septic shock *z7952* - Long term (current) use of systemic steroids *z79891* - Long term (current) use of opiate analgesic *z8249* - Family history of ischemic heart disease and other diseases of the circulatory system
Covid-19 (second half of 2020)	a - Certain infectious and parasitic diseases d - Blood and immune disorders e - Endocrine, nutritional and metabolic diseases f - Mental and behavioral disorders g - Diseases of the nervous system i - Diseases of the circulatory system j - Diseases of the respiratory system k - Diseases of the digestive system m - Diseases of the musculoskeletal system and connective tissue n - Diseases of the genitourinary system r - abnormal clinical findings z - Other health services	*e7800* - Pure hypercholesterolemia, unspecified *i252* - Old myocardial infarction *i959* - Hypotension, unspecified *k5900* - Constipation, unspecified *m1990* - Unspecified osteoarthritis, unspecified site *n1830* - Chronic kidney disease, stage 3 unspecified *z23* - Encounter for immunization *z370* - Single live birth *z880* - Allergy status to penicillin *z90710* - Acquired absence of both cervix and uterus *z955* - Presence of coronary angioplasty implant and graft

Fig. 4 shows the visualizations of ego networks for "z370" over the three six-month periods – before COVID-19 (second half of 2019), COVID-19 onset (first half of 2020), and COVID-19 (second half of 2020). "z370" is the ICD-10 code for a single live birth. Despite the fact that another ICD-10 code, "z372," represents both living babies (twins), "z370" has a considerably higher incidence in diagnoses connected to newborns and should be representative for the examination of COVID-19's influence on neonates.

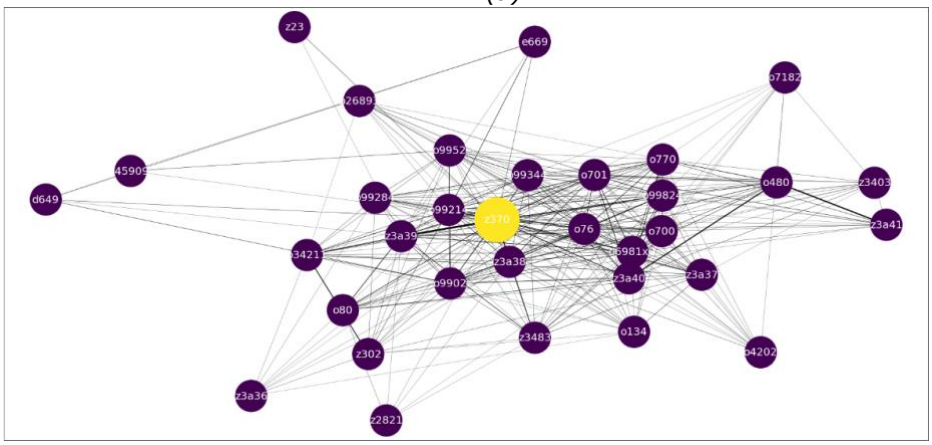
Fig. 4 (a) includes 32 nodes and 275 edges, implying that each diagnosis in this ego network had an average of 17.1875 co-occurring diagnoses in the second half of 2019, based on a weight bar of 0.05. The 32 diagnoses in this ego network may be classified into five groups (Table 3: Category). Among the five categories, "Pregnancy, childbirth and the puerperium," the most popular category of diagnoses that co-occurred with "z370," had 17 appearances, followed by "Other health services" with 12 appearances, and one appearance for "Blood and immune disorders," "Endocrine, nutritional and metabolic diseases," and "Diseases of the respiratory system." The average weight of the edges in Fig. 4 (a) is 0.1067.

In terms of the number of nodes, edges, average degree, and average weight of the edges in the ego networks, Fig. 4 (b) and Fig. 4 (a) have extremely comparable ego network structures. Fig. 4 (b) shows that the four variables have values of 33, 278, 16.8485, and 0.1062, respectively. According to the Table 3, in the first half of 2020, the diagnosis of o99334 - Smoking (tobacco) complicating childbirth, which was in the ego network of "z370" in the second half of 2019, was replaced by o26893 - Other specified pregnancy related conditions at the third trimester, and o7182 - Other

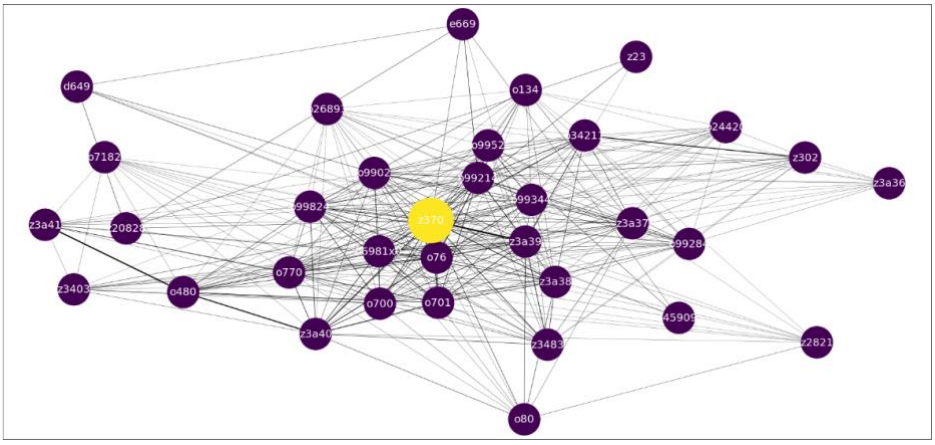
specified trauma to perineum and vulva.



(a)



(b)



(c)

Fig. 4. Ego networks of z370 - Single live birth for (a) 2019-02, (b) 2020-01, and (c) 2020-02.

Fig. 4 (c), the ego network of "z370" in the second half of 2020, contains 34 nodes, 308 edges, and an average degree of 18.1176. The average weight of edges in Figure 4 (c) is 0.1057. The ego networks of "z370" in the first and second half of 2020 have 32 common nodes. The diagnosis of *o4202* - Full-term premature rupture of membranes, onset of labor within 24 hours of rupture was in the ego network of "z370" in the first half of 2020, yet it was removed from the second half of 2020 with the weight bar of 0.05. Instead, as shown in Table 3, two new diagnoses joined the ego network: *o24420* - Gestational diabetes mellitus in childbirth, diet controlled, and *z20828* - Contact with and (suspected) exposure to other viral communicable diseases. As discussed in the section of the ego networks of "z20828," given the weight bar of 0.05, COVID-19 became one of the principal diagnoses linked to pregnancy and newborns. Besides that, the observations of Gestation diabetes mellitus (GDM) in the ego network of "z370" in the second half of 2020 also implied that COVID-19 might have had direct contributions to complicating the process of the pregnancy and newborn. According to an assessment of over 1000 pregnancies conducted at a leading UK hospital, women diagnosed with GDM during the COVID-19 pandemic were detected later in their pregnancy than those identified before the pandemic [21]. Therefore, if the women became pregnant and developed GDM during the COVID-19 onset period (first half of 2020), their diagnoses may not be identified until later in their pregnancy, resulting in the fact that "z370," "o24420," and "z20828" were linked in the same ego network in the second half of 2020 with a weight bar of 0.05.

Table 3. Information of the ego networks of "z370"

Time period	**Category**	**Unique diseases / new diagnosis from the previous period**
Before COVID-19 (second half of 2019)	d - Blood and immune disorders e- Endocrine, nutritional and metabolic diseases j - Diseases of the respiratory system o - Pregnancy, childbirth and the puerperium z - Other health services	*o99334* - Smoking (tobacco) complicating childbirth
COVID-19 onset (first half of 2020)	d - Blood and immune disorders e- Endocrine, nutritional and metabolic diseases j - Diseases of the respiratory system o - Pregnancy, childbirth and the puerperium z - Other health services	*o26893* - Other specified pregnancy related conditions at the third trimester *o7182* - Other specified trauma to perineum and vulva
COVID-19 (second half of 2020)	d - Blood and immune disorders e- Endocrine, nutritional and metabolic diseases j - Diseases of the respiratory system o - Pregnancy, childbirth and the puerperium z - Other health services	*o24420* - Gestational diabetes mellitus in childbirth, diet controlled *z20828* - Contact with and (suspected) exposure to other viral communicable diseases

The average degree and average weight of edges in the ego network of "z370" in the first half of 2020 decreased as the prevalence of "z370" decreased by 9.53 percent from the pre-COVID-19 period to the COVID-19 beginning period. While the average weights of the edges in the ego networks have remained steady across the three periods, the number of edges has increased by about 10.79 percent from the

first to the second half of 2020, suggesting that, despite the fact that the number of nodes associated to "z370" did not grow much throughout the pandemic, the number of complications of diseases risen dramatically.

Fig. 5 depicts three "z371" - single stillbirth's ego networks, each corresponding to a six-month period throughout the epidemic. These three graphs should aid in determining how illnesses connected to a single stillbirth have changed over time while under the effect of COVID-19.

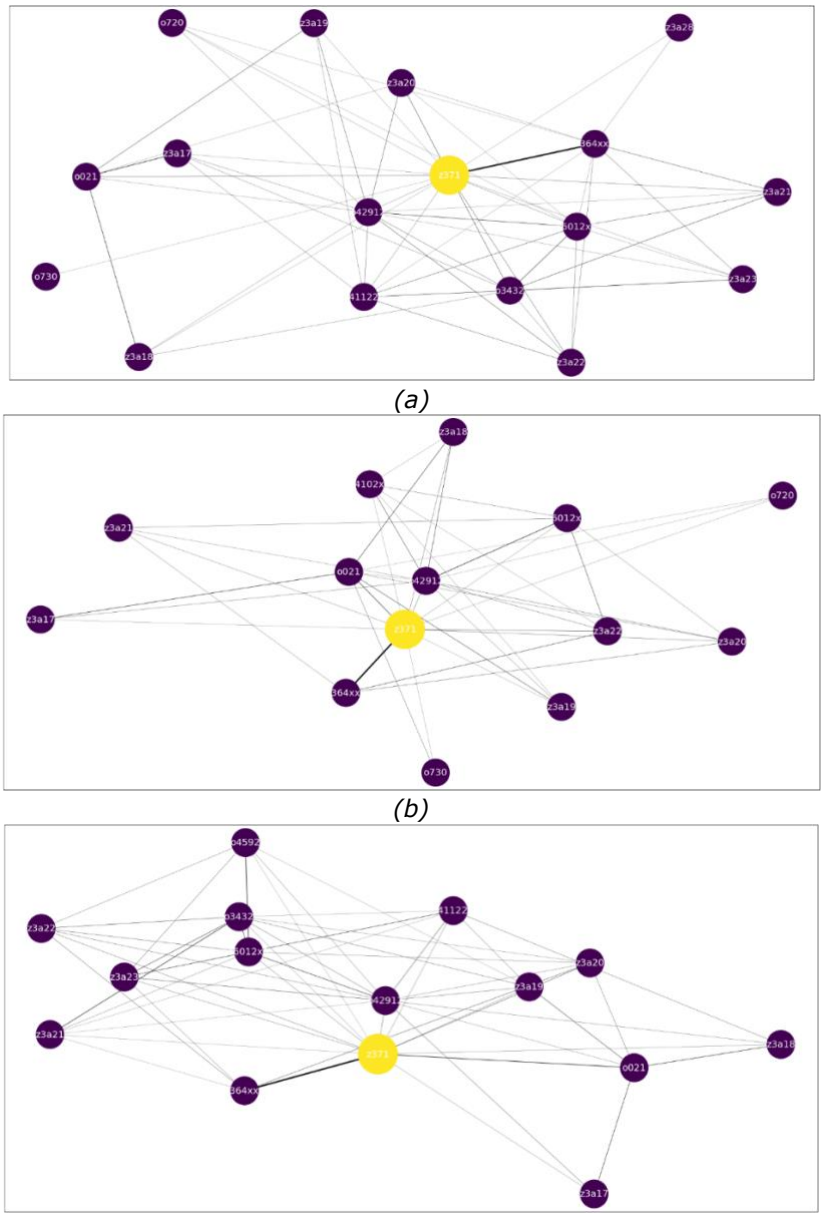


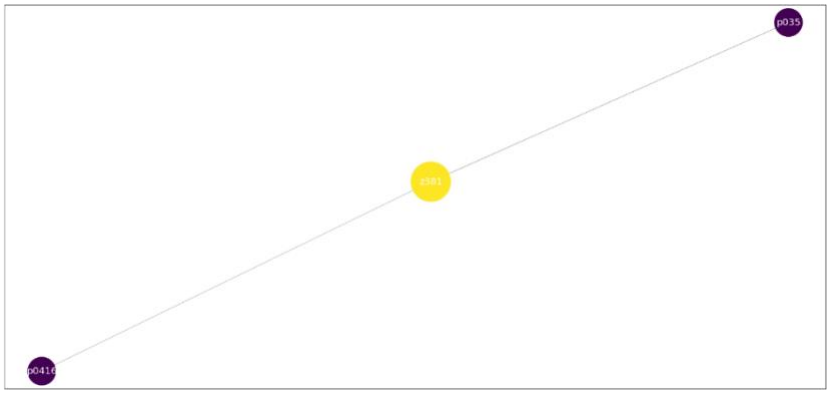
Fig. 5. Ego networks of z371 - Single still birth for (a) 2019-02, (b) 2020-01, and (c) 2020-02.

The ego network of "z371" in the second half of 2019 is shown in Fig. 5 (a),

with 17 nodes and 57 edges, and an average edge weight of 0.0972. Eight diagnoses in the ego network are categorized as "pregnancy, childbirth and the puerperium," while another nine diagnoses are classified as "Other health services." Fig. 5 (b) has 14 nodes and 39 edges with an average weight of edges of 0.1057 and has 13 common nodes with Fig. 5 (a). Fig. 5 (c) has 11 common nodes with Fig. 5 (b), an average weight of edges of 0.1081, 15 nodes and 54 edges. The prevalence and ego network complexity all decreased in the first half of 2020, and then increased in the second half of 2020, even passing the levels before the COVID-19 pandemic (second half of 2019), according to the column of "Unique diseases / new diagnosis from the previous period" in Table 4 and the single still birth graph in Fig. 2, with a weight bar of 0.05.

Table 4. Information of the ego networks of "z371"

Time period	**Category**	**Unique diseases / new diagnosis from the previous period**
Before COVID-19 (second half of 2019)	o - Pregnancy, childbirth and the puerperium z - Other health services	*o3432* - Maternal care for cervical incompetence, second trimester *o411220* - Chorioamnionitis, second trimester, not applicable or unspecified *z3a23* - 23 weeks gestation of pregnancy *z3a28* - 28 weeks gestation of pregnancy
COVID-19 onset (first half of 2020)	d - Blood and immune disorders e- Endocrine, nutritional and metabolic diseases j - Diseases of the respiratory system o - Pregnancy, childbirth and the puerperium z - Other health services	*o4102x0* - Oligohydramnios, second trimester, not applicable or unspecified
COVID-19 (second half of 2020)	d - Blood and immune disorders e- Endocrine, nutritional and metabolic diseases j - Diseases of the respiratory system o - Pregnancy, childbirth and the puerperium z - Other health services	*o3432* - Maternal care for cervical incompetence, second trimester *o411220* - Chorioamnionitis, second trimester, not applicable or unspecified *o4592* - Premature separation of placenta, unspecified, second trimester *z3a23* - 23 weeks gestation of pregnancy



(a)

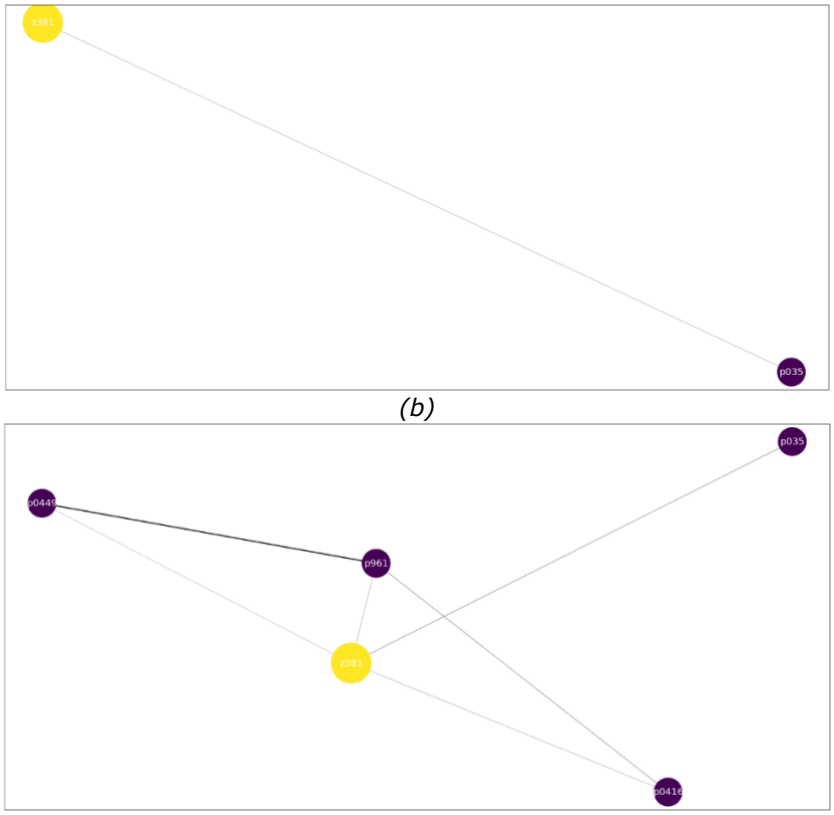


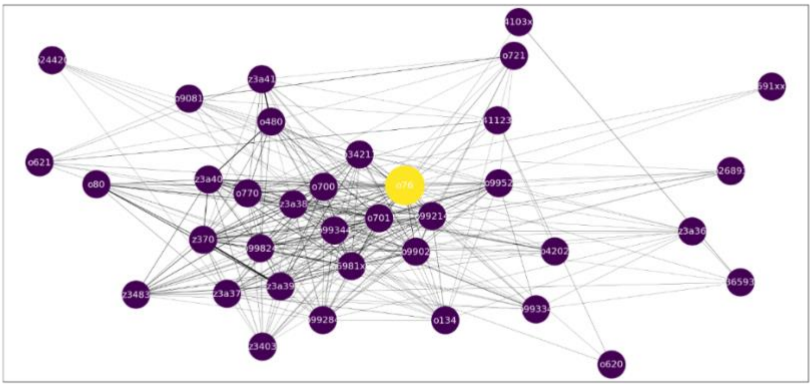
Fig. 6. Ego networks of z381 - Single liveborn infant, born outside hospital for (a) 2019-02, (b) 2020-01, and (c) 2020-02

The ego network for ICD-10 code z381 should reveal how the disease network altered for the infant who was born outside of the hospital. Throughout the three periods, "z381's" ego networks are sparser than the preceding ones. Furthermore, according to the column "Unique disease / new diagnosis from the previous period" in Table 5, diagnoses connected to maternal usage of addictive substances were found in the ego network during the COVID-19 epidemic (second half of 2020), implying that the maternal health problem outside of hospitals should be further investigated.

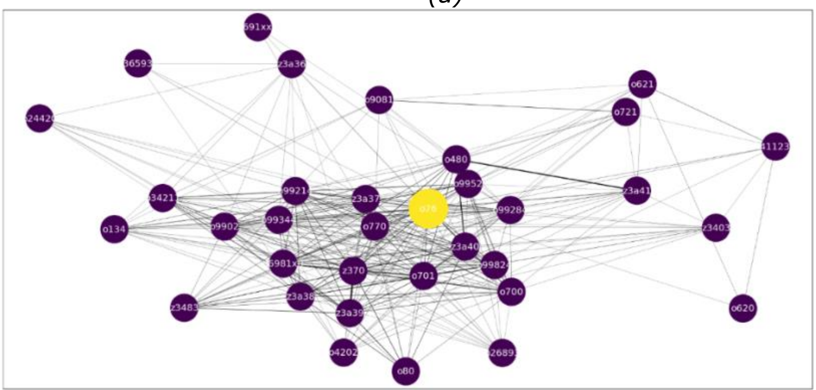
Table 5. Information of the ego networks of "z381"

Time period	**Category**	**Unique diseases / new diagnosis from the previous period**
Before Covid-19 (second half of 2019)	p - Certain conditions originating in the perinatal period	*p0416* - Newborn affected by maternal use of amphetamines
Covid-19 onset (first half of 2020)	p - Certain conditions originating in the perinatal period	None
Covid-19 (second half of 2020)	p - Certain conditions originating in the perinatal period	*p0416* - Newborn affected by maternal use of amphetamines *p0449* - Newborn affected by maternal use of other drugs of addiction *p961* - Neonatal withdrawal symptoms from maternal use of drugs of addiction

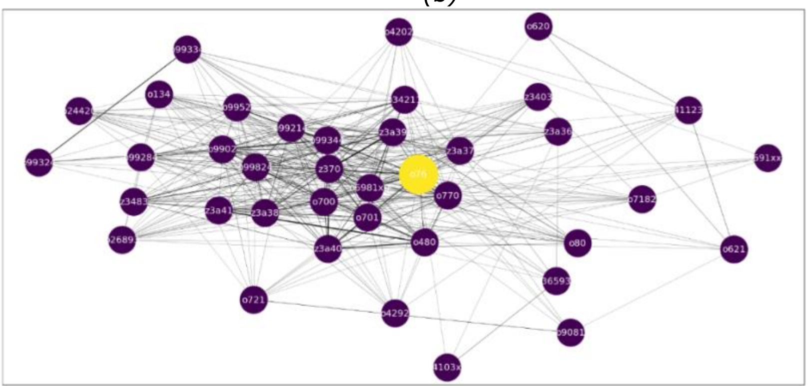
The diagnosis of o76 has the largest number in abnormal, complicated deliveries, which should help us understand the detrimental influence of the pandemic environment on the delivery process and maternal health complications. In Figure 7, the average edge weights in the ego networks are 0.1006, 0.1029, and 0.0989, respectively. Moreover, for the graphs in Fig. 7, the average degrees are 16.6111, 16.7647, and 17.8974, respectively, leading to a positive correlation with the number of COVID-19’s confirmed cases throughout the three six-month periods.



(a)



(b)



(c)

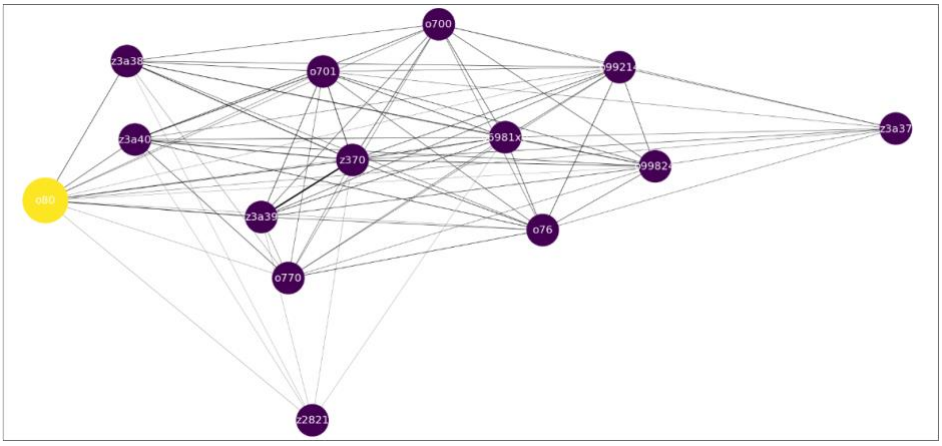
Fig. 7. Ego networks of o76 - Abnormality in fetal heart rate and rhythm complicating labor and delivery for (a) 2019-02, (b) 2020-01, and (c) 2020-02.

To put it another way, as the pandemic progressed, the average number of co-occurrences of each disease in the ego network of "o76" increased. As the pandemic spread (from the second half of 2019 to the first half of 2020), the weights of the links connecting "o76," "o4103x0," and "o99334" fell below 0.05, according to Table 6. However, by the second half of 2020, two diagnoses, "o4103x0" and "o99334," had rejoined the ego networks of "o76," and three new diagnoses, "o4292," "o7182," and "o99324," had emerged as well.

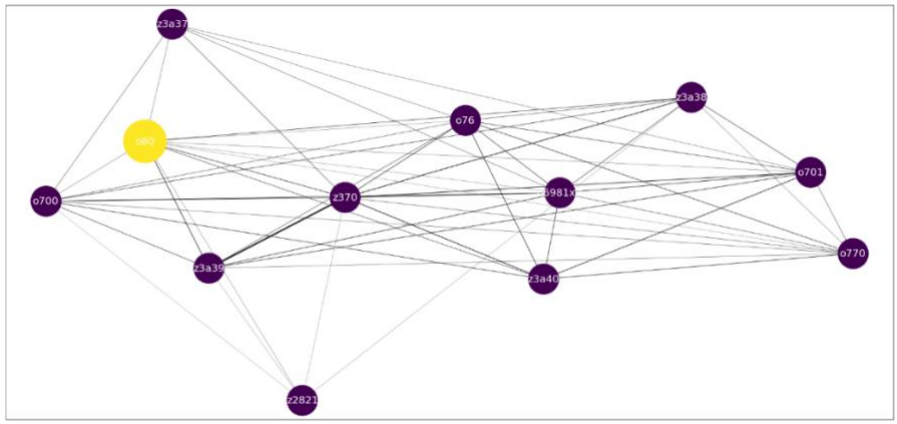
Table 6. Information of the ego networks of "o76"

Time period	**Category**	**Unique diseases / new diagnosis from the previous period**
Before Covid-19 (second half of 2019)	o – Pregnancy, childbirth and the puerperium z – Other health services	*o4103x0* – Oligohydramnios, third trimester, not applicable or unspecified *o99334* – Smoking (tobacco) complicating childbirth
Covid-19 onset (first half of 2020)	o – Pregnancy, childbirth and the puerperium z – Other health services	*None*
Covid-19 (second half of 2020)	o – Pregnancy, childbirth and the puerperium z – Other health services	*o4103x0* – Oligohydramnios, third trimester, not applicable or unspecified *o4292* – Full-term premature rupture of membranes, unspecified as to length of time between rupture and onset of labor *o7182* – Other specified trauma to perineum and vulva *o99324* – Drug use complicating childbirth *o99334* – Smoking (tobacco) complicating childbirth

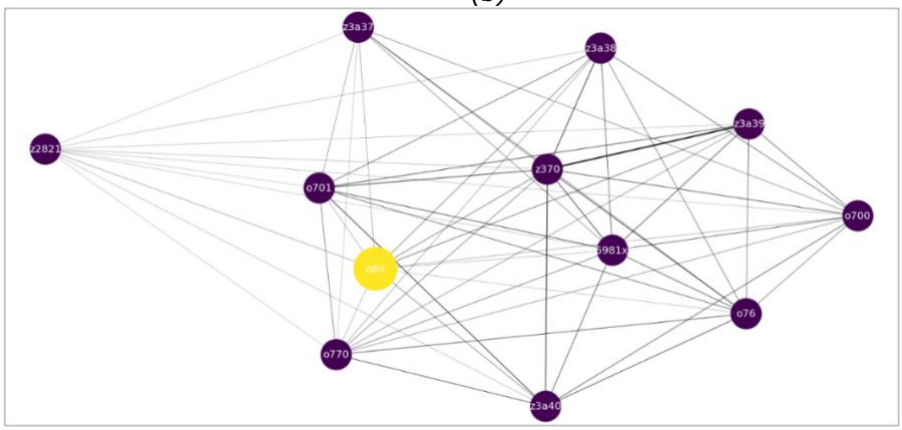
According to Fig. 8, the numbers of nodes are 14, 12, 12 for each sub graph. And the average degrees are 10.8571, 8.6667, and 9.6667, respectively.



(a)



(b)



(c)

Fig. 8. Ego networks of o80 - Encounter for full-term uncomplicated delivery for (a) 2019-02, (b) 2020-01, and (c) 2020-02.

And the average weight of the edges in each ego network are 0.1457, 0.1512, and 0.1438. According to Table 7 and the prevalence plot of "o80" in Fig. 2, there were no new diagnoses in the ego networks of "o80" in 2020 as the number of uncomplicated deliveries decreased significantly in the first and second half of 2020.

Table 7. Information of the ego networks of "o80"

Time period	**Category**	**Unique diseases / new diagnosis from the previous period**
Before Covid-19 (second half of 2019)	o - Pregnancy, childbirth and the puerperium z - Other health services	*o99214* - Obesity complicating childbirth *o99824* - Streptococcus B carrier state complicating childbirth
Covid-19 onset (first half of 2020)	o - Pregnancy, childbirth and the puerperium z - Other health services	*None*
Covid-19 (second half of 2020)	o - Pregnancy, childbirth and the puerperium z - Other health services	*None*

5. Conclusion

The COVID-19 epidemic has wreaked havoc on human society, and its impact on health and wellbeing is still not fully known. Understanding and minimizing the effects of the pandemic on pregnancy and childbirth is an important problem to be addressed. This study has focused on analyzing the variations in hospitalization and treatment of expectant mothers and newborn care in broad groups. Analyzing the changes of the disease networks and the variations in the ego networks of specific diagnoses using the EHR data from Arizona Department of Health Services over three six-month periods- before COVID-19 (second half of 2019), COVID-19 onset (first half of 2020), and COVID-19 (second half of 2020) can help understand how the pregnancy and newborns were affected under the circumstance of the pandemic. Except for "o76," the average degrees of all other diagnoses studied in this study decreased in the first half of 2020, implying that many regular diagnoses in disease networks were underreported in the first half of 2020 because most hospitals were not prepared to deal with the crisis. In addition, compared to the first half of 2020, the number of new co-occurring diagnoses grew substantially in the second half of 2020 except for "o80". The study of ego networks of "z381" and "o76" also reveals that as the pandemic environment got harsh in the second half of 2020, the use of drugs and drugs of addiction contributed to complicating the delivery process. In the second half of 2020, pregnant women and infants were more prone to develop complications or be diagnosed with diseases that were rare before COVID-19. The findings suggest that COVID-19 may have had a direct or indirect influence on hospital visits with pregnant women and infants during the pandemic, indicating that more research into the cause of such population shifts is needed.

References

1. Agusto, F. B. *et al.* Impact of public sentiments on the transmission of COVID-19 across a geographical gradient. *medRxiv* 2021.01.29.21250655 (2021) doi:10.1101/2021.01.29.21250655.
2. Iwendi, C. *et al.* COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health* **8**, (2020).
3. Yadaw, A. S. *et al.* Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digital Health* **2**, (2020).
4. Jamieson, D. J. & Rasmussen, S. A. Protecting Pregnant Women and Their Infants From COVID-19: Clues From Maternal Viral Loads, Antibody Responses, and Placentas. *JAMA network open* vol. 3 (2020).
5. Gharoie Ahangar, R., Pavur, R., Fathi, M. & Shaik, A. Estimation and demographic analysis of COVID-19 infections with respect to weather factors in Europe. *Journal of Business Analytics* **3**, (2020).
6. Herbrich, R., Rastogi, R. & Vollgraf, R. CRISP: A probabilistic model for individual-level COVID-19 infection risk estimation based on contact data. *arXiv* (2020).
7. Gephi. Gephi Tutorial Quick Start. *Gephi* (2010).
8. Barabasi, A. & Frangos, J. *Linked: The New Science Of Networks Science*. (Perseus Books Group, 2014).
9. Harenberg, S. *et al.* Community detection in large-scale networks: a survey and empirical evaluation. *Wiley Interdisciplinary Reviews: Computational Statistics* **6**, 426–439 (2014).
10. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, (2002).
11. Nordhausen, K. Statistical Analysis of Network Data with R. *International Statistical Review* **83**, (2015).
12. Dai, Y., Wang, S., Xiong, N. N. & Guo, W. A survey on knowledge graph embedding: Approaches, applications and benchmarks. *Electronics (Switzerland)* **9**, (2020).
13. Srinivasan, K., Currim, F. & Ram, S. Predicting High Cost Patients at Point of Admission using Network Science. *IEEE Journal of Biomedical and Health Informatics* **22**, 1970–1977 (2018).

14. Hidalgo, C. A., Blumm, N., Barabási, A. L. & Christakis, N. A. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Computational Biology* **5**, (2009).
15. Steinhaeuser, K. & Chawla, N. v. A Network-Based Approach to Understanding and Predicting Diseases. in *Social Computing and Behavioral Modeling* 1–8 (Springer US, 2009).
16. Healy, C. M. COVID-19 in Pregnant Women and Their Newborn Infants. *JAMA Pediatrics* (2021) doi:10.1001/jamapediatrics.2021.1046.
17. Schwartz, D. A. An analysis of 38 pregnant women with COVID-19, their newborn infants, and maternal-fetal transmission of SARS-CoV-2: Maternal coronavirus infections and pregnancy outcomes. *Archives of Pathology and Laboratory Medicine* **144**, (2020).
18. CDC. Breastfeeding and Caring for Newborns. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/pregnancy-breastfeeding.html (2021).
19. Pollard, T. J. *et al.* The eICU collaborative research database, a freely available multi-center database for critical care research. *Scientific Data* (2018) doi:10.1038/sdata.2018.178.
20. Reyna, M. A. *et al.* Pathway and network analysis of more than 2500 whole cancer genomes. *Nature Communications* **11**, (2020).
21. Davenport, L. COVID-19 May Have Worsened GDM Pregnancy Outcomes. *Medscape* https://www.medscape.com/viewarticle/950009 (2021).

The impact of big data on innovation and value generation in pharmaceutical sales and marketing

Antonio Pesqueira[1][0000-0003-1530-6451]

[1] Bavarian Nordic A/S, Grafenauweg 8, 6301 Zug, Switzerland

https://doi.org/10.33847/2686-8296.3.2_3

Received 13.10.2021/Revised 07.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. Using Big Data in the pharmaceutical industry is a relatively new technology, and the benefits and applications are yet to be understood. There are some cases currently being piloted, but others have already been adopted by some pharmaceutical organizations, proving the unmet need in a field that is still in its infancy.

This paper aims to understand how and if Big Data can contribute to commercial innovation, as well as future trends, investment opportunities.

Participants from 26 pharmaceutical companies participated in different focus groups where topics were grouped by individuals and evaluation areas were discussed to discover any potential connections between Big Data and Innovation in commercial pharmaceutical environments. This study used the collected data to analyze and draw conclusions about how many life sciences leaders and professionals already know about Big Data and are identifying examples and processes where Big data is supporting and generating innovation.

In addition, we were able to understand that the industry is already comfortable with Big Data, and there were some very accurate research results regarding the most pertinent application fields and key considerations moving forward.

Using the network analysis findings and the relationships and connections explained by respondents, we can reveal how Big Data and innovation are interconnected.

Keywords: Big Data, Digital, Pharmaceutical Industry, Focus Group, Commercial

1. Introduction

Companies in the pharmaceutical industry face intense competition from a variety of sources in the operating markets. To overcome the current difficulties and barriers, it may be necessary to seek new methods of Research and Development (R&D) as well as new operational mechanisms to achieve economic scale and enhance the effectiveness of organizational investments [1].

In most cases, innovation comes from outside an organization rather than from its core, which is why Big Data (BD) is an example since historically these technologies are developed by outside forces. Pharma companies are currently searching for new disruptive methods for reducing business risks and preparing commercial models for a more technologically advanced approach by adopting innovative business models [2]. In addition to the significant challenges that pharmaceutical companies continue to face because of a crescendo of increasing competition, decreasing revenues, expiring product patents, and limiting access to physicians, competitors still maintain positive momentum, with strong predictions of continued success [3].

The complexity of pharmaceutical effectiveness operations has been influenced by constant pressure and unrelenting regulatory scrutiny, with severe impacts from health care reforms and fewer drugs approved around the world, but specifically in Europe where escalating prices force governments there to impose fair trade conditions and price cuts. Therefore, pharmaceutical companies today are not only

expected to continue to leverage and grow geographic scale, establish strong market positions, adopt new technologies, and properly adapt various processes.

In the pharmaceutical industry, BD has led to an increase in productivity and efficiency, especially in business processes, as reported in several different research studies published recently. Through the development of tasks and commercial design models, commercial leaders can adjust the workflow and tasks accordingly to meet a particular need or problem in the market. It provides insight into patterns of behavior and predictions of what critical commercial functions like supply chain management, market access, market research, and commercial excellence need to be optimized [4].

In the pharmaceutical industry, BD analytics is utilized to predict which products and commercial events will lead to new engagement models with healthcare practitioners and decision-makers, and accordingly, to analyze and place new competitive tactics or brand plans. These are simple examples of how BD can be used, though the potential of this industry has yet to be realized. It can be used for market research activities, interactions with patients, or even support services related to a company's products and services.

In this research study, we present a brief study of innovation fundamentals and theories and contribute to the review of basic ideas related to the innovation process in the pharmaceutical industry, and we discuss the main issues for potentiating BD in the pharmaceutical industry.

It will therefore assess current BD applications in commercial pharmaceutical functions, trends in the future, and key characteristics of existing applications. It will also assess possible connections and relationships which suggest BD drives innovation and can be considered as a valuable innovation source.

The study included 26 participants from various industries and business functions and examines the implications of these changes regardless of whether BD technology is used within businesses.

We will analyze the focus group findings closely and support the research of BD as an innovation source for pharmaceutical processes and areas considering the current adoption and application of this technology.

The data collected in this focus group survey came from interviews and discussions conducted from June 2021 to July 2021, with the objective of better understanding how the life sciences industry views BD technology in general.

Following are the remaining sections of this document. A literature review of how it relates to the topic can be found in section 2. Next, a discussion regarding the methodology of a focus group is presented for assessing BD-based solutions for commercial pharmaceutical functions, followed by a summary of the findings.

2. Literature Review

2.1. Big Data in Pharmaceuticals

The pharma industry handles some of the most complex and time-consuming projects in the world, where a considerable number of resources, submissions, and compliance approvals or validations are used in a pharma company that generates large amounts of data.

Big Data (BD) can analyze extremely large sets of data to find hidden patterns and trends that can be turned into actionable insights. Previously unknown correlations can be used to create informed decisions for the business, and it provides one of the most valuable pieces of technology in the pharmaceutical industry.

Historical data finds patterns and probabilities to commercial strategy models and projects towards success and avoids any risks or pitfalls.

By pulling data from previous commercial projects or business cases that have been collected from historical decision processes or business initiatives, BD can reveal the extent of how these business cases were used and lead to insights into what is worth investing and what is frequently used and may be worth investing in. Hence, it is essential to take all the necessary money-saving measures to manage a BD project within the pharmaceutical commercial space effectively [5].

In the pharmaceutical industry, BD refers to the enormous amount of data and information which has been stored in advanced computing systems and connected with different functions but especially commercial.

While understanding the value of BD continues to remain a challenge, other practical challenges, including funding and return on investment and skills, continue to remain at the forefront for several different industries that are adopting BD without pharmaceutical being an exception. Generally, most pharmaceutical organizations have several goals for adopting BD projects, but few have already well-established frameworks or systematic processes within commercial functions. While the primary goal for most organizations is to enhance customer experience, other goals include cost reduction, better-targeted marketing, and making existing processes more efficient. In recent times, data breaches have also made enhanced security an important goal that BD projects seek to incorporate. By leveraging BD, businesses can process huge amounts of data at unprecedented speeds and will be able to focus on their primary activities rather than on handling the data [6].

Using BD tools, every data-related task can be managed more efficiently, like database management, reporting, etc. There are many challenges that companies must face, such as low-quality resources, unpredictable market conditions, and regulations, where data can be used to create a virtual scene with the aid of simulation tools and software to find every possible outcome.

Several of the practices from the market access departments require the use of BD to understand the impact of pricing and economic dossiers submitted to the local regulators and authorities.

Clinical trial projects are monitored in real-time by BD, which enables companies to manage patient involvement and schedule the phases of many trials efficiently using large amounts of data. BD analytics are very helpful in providing valuable insights to avoid or solve potential problems and finding new areas to improve cost and efficiency [7].

In the rare disease industry, for instance, BD analytics is used to predict the risks associated with new research discoveries, publications, or medical interactions. By performing queries and searches on the data, facilitates the process of making better decisions and reduces project risks by predicting in simulating environments the responses and preferences of Key Opinion Leaders (KOLs) and Healthcare Professionals (HCPs). BD in pharmaceutical products sales and supply chain activities is used to analyze high-frequency products transactions, to support sales and marketing decision-making, to measure sentiment, or to apply predictive analytics to processes such as segmentation and sales force targeting [8].

A large amount of BD is also used by the industry for risk analytics, such as parallel trading, product supply, enterprise risk management, forecasting, and tendering.

Data integration services such as PubMed for medical publications and clinical trials involvement have been used by commercial insights teams to build visual data that helps identify KOLs faster and analyze healthcare information more efficiently,

which is used to track the spread of chronic diseases and promote clinical products. Another use of BD in pharma is to measure the effectiveness of pharmaceutical commercial teams when engaging relevant HCPs or KOLs [9].

Several other BD applications, including revenue management, technological enhancement, logistics, and competitive advantage or forecast related to sales and supply chain, have been developed using massive data from digital channels and social media. A 360-degree view of customers can already be attained by several pharmaceutical companies through BD, which allows them to optimize marketing, increase sales, and improve customer service, as well as strengthen counterfeit drug prevention and cybersecurity protection by identifying suspicious transactions and security threats. Financially, several companies are already using BD for improving business forecasts and processes, optimizing product pricing, and increasing operational efficiency, and analyzing text, images, and audio to identify patterns and better match content to medical audiences and plan better product launches. Some BD challenges in the commercial process include gathering, analyzing, and utilizing real-time insights from HCPs or KOLs, or understanding patterns of real-time, media content usage from field sales teams [10].

In this industry, commercial organizations analyze customer data along with behavioral data to create personalized customer profiles to cater to different target audiences, recommend medical content on-demand, or even measure the performance of digital marketing campaigns [11].

It is a significant technical challenge for the industry to integrate BD from different sources and vendors into platforms that were designed for market research and competitive intelligence.

Data derived from social media, medical societies, events, and clinical trial participants have also been used in the industry to analyze and predict customer behavior to provide transparent and better product engagement plans.

Business operations, products, and services can be analyzed using BD, such as engines and other project problems, enabling businesses to recognize and fix problems as soon as they arise, so that the business and the project can work efficiently. Aspects of the business are being modernized to modernize the market and the product development phases. This includes everything from clinical development to real-world evidence to commercialization [12].

2.2. Innovation Concepts in Commercial Pharmaceuticals

Organization studies in organizational microsystems of innovation addressed a different perspective from the innovation theory literature, namely that innovations happen mainly within a national system of innovation. Process innovation, such as implementing new methods and policies within an organization, as well as developing new business objectives and strategies, is organizational innovation.

Literature shows that the concept of innovation is very complex, which makes it difficult to have a single definition, were in some other pharmaceutical context and according to our initial literature review from the industry typology also differentiates three types of innovation: processes, product or services, and strategy or business [13].

Alternatively, there are various classifications of innovation in the literature. Pharmaceutical innovations are structured to occupy three areas: enhancing and expanding the range and scope of products and services, redesigning the production

and regulatory processes, and implementing supply chain management systems; and introducing changes to management, work organization, and employee qualifications [14].

The pharmaceutical industry's survival depends on an increasing level of competitiveness, and organizational innovation can play an important role, especially since organizations invest so consistently in innovation strategies. Answering this question is a fundamental and complex dilemma since innovation is not explicitly linked to competitiveness and investing in technology or people raises questions about the organizations' long-term survival [15].

The last few years have seen some new concepts emerge, as well as an increase in corporate and academic knowledge from benefits analysis to innovations. It is not only a concept of great relevance to business but also a source of influence over the dynamics and models of business as a whole [16].

There are different levels of innovation such as incremental improvements, radical changes, or a total paradigm shift. By incremental innovation, we mean improvements in products and procedures that upgrade the quality and can still be cost-effective and productive. As a radical, the discontinuous result of different business capabilities or functional practices is referred to. It means that large economic changes lead to completely new business strategies such as new products, processes, changes in economic and social structure, and the behavior of economic leaders [17].

Branches, clients, and market niches can be created or destroyed by radical innovations, leading to new skills, abilities, and knowledge within the same market.

In contrast to incremental innovation, radical innovation usually follows along with a technology push, which means that new concepts and paradigms are created inside the corporation, often by scientists or engineers dedicated to research and development. On the other hand, radical innovations are rarely altered or altered in their development stages, primarily to protect the overall concept, which is too delicate to be altered or changed during this phase [18].

We are not always successful in bringing radical innovations to the market, regardless of their quality, because extreme innovation requires an enormous amount of effort. Innovative products & technological innovations; innovative procedures; innovative organizations; and marketing innovations are the different types of innovation. A fourth way to perceive innovation is to see it in terms of management innovation, strategic innovation, product innovation, or operational innovation. As a result of each level of innovation, competitive value outcomes are asymmetrical, but not all levels result in competitive advantage [19].

An organization's internal and external sources can be formal or informal. It is therefore crucial that pharma companies arrange their processes and activities to identify market trends and opportunities to remain competitive and be as efficient as possible. In commercial and other functions like market access, the ability to develop tools to track competitors' competitive advantage, their clients, and social contexts surrounding them is extremely relevant [20].

Organization location, region, sector, organizational dimension, globalization level, and relation to environments with catalyzing abilities influence innovation outcomes. A company's innovation practices influence what data is available and how it is utilized, as well as the client, supplier, and partner relationships [21].

The sources of innovation are variable, but the most relevant are patient product and supplier elements. There is a need to emphasize the useful sources of innovation

since each source plays a part inside the innovation process and relates to the future goals of the innovation. An organization or the people who will directly benefit from a product's innovations, procedures, or services are the functional sources of innovation. Organizations and people vary according to the innovation's analysis based on their functional relationship with innovators, users, and innovations. These organizations and people do not sit passively by, they actively participate in innovation [22].

The solutions to these problems can be explained using multidisciplinary collaborations and cooperation, as well as advanced commercial models leveraging key account management. The risk-sharing activity models in commercials are the main characteristics of these pharmaceutical companies [23].

2.3. Big Data, Digital, and Innovation

Keeping up with the regular market, technology, and competitiveness requires innovation in pharmaceutical commercial functions. New products, services, and systems arise through these mechanisms. Pharmaceutical products are the backbone of any successful organization. Creating new business models and changing the lives of many patients while also growing your business gives you a competitive advantage [24].

The use of new technology such as data catalog software, along with such innovations, will also help pharmaceutical companies generate higher profits and grow more effectively.

To differentiate, adapt to change, and seek the greatest return on investment, many organizations support the development of themselves and their businesses. The use of Big Data (BD) can also be used to develop procedures, products, and services that are better suited for understanding the competitive environment. Through the appropriate perception of the surrounding context, better knowledge, and understanding of the various market influences, BD is used as an objective competitive edge for many current organizations [25].

As a result of several corporate systems appearing in the past few years, decision centers have been able to analyze data, model market behavior, and retrieve, collect, and process information. Across all those questions, BD can provide answers, and in addition, it is completing the decision-making center, almost simultaneously with the market and its external variables [26].

The use of BD can therefore be considered as an innovation in-sight tool, as well as a tool to observe the market, analyze customers' strategic behavior, compete with pharmaceuticals, and evaluate all market necessities [27].

For innovation to reach its full potential, the market needs a need to obtain this product or final service so that it can be evaluated, monitored, and controlled. As an opponent develops their unique capabilities and strength in the market, BD can help a pharmaceutical organization better understand the market. Another advantage of using BD for innovation in products and services is the ability to gather and analyze perceptions and opinions of HCPs, KOLs, or patients regarding specific products and services available. It allows for information to be collected through BD in response to market-specific innovations, which will directly impact how innovation is made and developed [1].

Rare pharmaceutical companies are bringing commercial and R&D teams together to build internal synergies and provide quality feedback during data

collection. As companies gather feedback, they can generate better products that help them to remain competitive. Some organizations have developed sophisticated BD capabilities that enable them to keep an eye on opportunities and threats to gain a greater understanding of the business environment, to enable them to enhance their strategic planning [28].

This new knowledge offered by BD tends to help the internal processes in terms of competitive advantage because of a more complete understanding of variables surrounding the organization.

As BD is focused on gathering and analyzing information for the benefit of leading corporate decision-makers, it allows for the creation of strategies and solutions at critical stages of the innovation process, in tandem with the financial and strategic implications of the investigations and development assumptions. However, organizations can create internal market necessities through innovative products or services, another necessity still exists due to the competitive nature of the market. This is a necessity because of the competition with the competitors as well as the patients, KOLs, HCPs, or any other product decision-maker needs [4].

3. Data and Methodology

The purpose of this section is to share an overview of the selected methodology as well as key discussion topics and highlights from all of the research selections. This study examines Big Data (BD) as an innovation source and influencing force to current pharmaceutical commercial challenges, such as utilizing resources, data, decision-making capabilities, and agility in commercial procedures with efficient utilization of resources.

As such, the selected research question aims to elucidate commercialization processes and capabilities within the pharmaceutical industry that are heavily influenced by BD processes, processes, and business cases. As well as addressing the above research objective, this paper introduces a second goal: which BD techniques and models may influence commercial pharma projects?

As a result of the focus group conducted in June and July of 2021, several different questions were asked to gather all the relevant back-bone knowledge and foundation for further understanding of all answers and nuances related to the already presented research questions.

A wide variety of commercial positions levels were represented in our focus groups, including vice presidents, global and local heads, senior directors, managers, and business analysts via Microsoft Teams. Depending on the role participants played in the study, their expertise and knowledge of the topics in question, their industry experience, and their availability to participate, focus group participants were selected. Having an experienced panel of people from the pharmaceutical industry with knowledge of the selected topics and experience in the commercial realm was one of the main objectives [29].

Observations and conclusions from the literature review were used to develop the focus groups study design, and a subset of the participants and experts in a selected area were consulted about the overall design, including sample questions and topics drawn from the literature review.

Participants were divided into smaller groups with similar titles and responsibilities for four remote digital sessions. Each session lasted 60-90 minutes. PowerPoint slides served as a guide to the interviews and facilitated the sessions.

Three key topic areas formed the basis of the focus group study. The study included six related focus group questions in total.

During this focus group study, there were some risks of discovering sensitive and confidential information, including commercial confidential practices, key performance indicators, or specific metrics that might be deemed competitively sensitive.

To ensure the confidentiality of the participants and working organizations, during the focus groups it was decided that all diseases, products, and participants' names would not be referred to and therefore retracted.

Below is a table showing the research questions, key purposes, and data collection methods. It also includes the key objectives and investigations topics related to each question.

Table 5. Description and overview of all research objectives, research questions and presented research hypothesis that acted as a key structure strategy to the research paper.

Research objective	Research questions	Research hypothesis
Identify the differences between companies employing BD strategies and companies without such strategies and assess any possible relationship with Innovation.	What are some of the current applications, challenges, and future opportunities of BD in the pharmaceutical commercial space, and how it connects with innovation?	As companies adopt BD processes, current and future applications will generate innovation?
Identify which type of pharmaceutical companies are implementing BD applications and which case studies have been done between them, as well as what types of future applications are currently being researched.	In terms of case studies and opportunities, how is BD being assessed in the current times and how will it evolve in the future?	Companies are becoming more aware of the value and impact of BD on the overall business operations and strategy as future applications and opportunities expand.

We developed a list of six questions to guide the discussions from the focus group sessions, with participants receiving the questions in advance of their scheduled sessions to have a better preparation and to support the participants in terms of reflecting and voicing their opinions beforehand. Each group had a different composition based on their job duties and titles, so their responses (see table 2) were influenced by different perspectives.

Table 2. Representative table with the focus group questions and presented areas where the focus group questions were part of

Area	Questions
Innovation	Concerning Big Data programs and activities, what are the most important organizational innovation characteristics? Can you identify any areas of connection or relationship between Big Data and Innovation?
	Can you identify any areas where Big Data and innovation are related, and where has Big Data invested in innovation generation in terms of business development?
Processes and strategy	What are your strategies for creating awareness of the value of business development in your organization/department? Would you be able to give examples of when both concepts were used to generate value?
	Uses your organization Big Data metrics to measure digital and business development?
Technology and data	For Big Data, what tools and systems are utilized?
	How do your Commercial and Big Data processes connect to the data sets, data points, and sources?

We will discuss the responses to the above questions within the following sub-section, including key insights from the focus group interviews, as well as topics for discussion.

4. Focus Group Results and Discussion

Generally, the interview participants rated each interview as having a high level of relevance, as well as a good level of commitment and dedication. The participants also noted that pre-reading the questions was extremely helpful and important for the preparation of the interview. On a more abstract level, several participants referred to the idea of having a more well-developed digital and Big Data (BD) organizational culture and overall data governance capability with a deep understanding of the various challenges faced by this industry in making the most from digital and commercial initiatives.

The focus groups provided us with a clear understanding that most of the participants' companies employ business development in commercial to acquire HCPs insights, execute effective virtual engagement events, and access stakeholder information.

We learned about key HCP engagement tactics through focus group discussions, including congresses and publications and guidelines awareness, disease education sessions, and Key Account Management (KAM) strategy plans.

As well as these four areas, other discussion areas where business development already leads to innovation practice and business value emerged from the discussion. These will be reflected in the following bullet points.

4.1. Leadership and Cloud-Based Solutions

BD efforts and available budgets in departments are heavily influenced by top management. As a result of understanding business challenges and business problems, actionable insights are generated about customers' behavior, healthcare systems, and other stakeholders.

To manage and collect BD processes and activities, therefore, it is crucial to establish a centralized, yet scalable, setting.

In terms of BD process management, cloud-based services have yet to appear and having the ability to control computing and resource allocation in the future will be critical.

By utilizing cloud storage services for BD management, commercial data can be integrated and processed more efficiently. Therefore, commercial systems are not currently capable of handling large amounts of data where it is a clear case of several companies struggling with their IT/IS department's capacity to support commercial needs, as well as implement master data management (MDM) platforms and data governance protocols.

In focus group interviews, scalability was likely one of the main topics discussed. Traditional monitoring remains popular despite low technology capabilities. Several other parameters were also discussed that are considered before BD programs are selected: price for included features, compliance with data protection laws, regular backups and data recovery, and ease of integration and migration.

4.2. Clinical Data

Some participants pointed out the importance of BD for biomedical and cell and gene therapies (CGT) data, and how complex and diverse it can be because it represents what we know about a human organism, which is composed of cells with nuclei containing chromosomes. Commercial functions in biomedical and CGT are a key focus of some of the participants, who usually support the administration of medications to patients by healthcare organizations like hospitals. This requires the collection of different data sources from different HCPs, to support the fundamental shift in focus from relying on HCPs as prescribers to having hospitals as partners.

Clinical and biostatistics departments store complex molecule or sequence data sets and different methods (e.g., single-pass sequence, multiple-fold repetition) of a sequence are fundamental to understanding Real World Evidence (RWE) and maximizing strategies for scientific exchange meetings, developing digital assets, and developing a brand.

While data is the simplest component of knowledge and information, information is the means of communicating data to others. Computers can process data and turn it into information, and in a commercial context, knowledge can be acquired through formal or informal analyses of data and information that can be employed in a variety of commercial activities. Commercial functions, therefore, have an established sequence for storing and processing data, starting with raw data, extracting information, and synthesizing new knowledge, where a lot of data can create a small amount of knowledge because it is more condensed and synthetic. The knowledge contained in books and articles is often actionable, allowing people to understand how classification segmentation of relevant audiences and commercial targets differs depending on which technique is used.

The complexity level of segmentation frameworks keeps increasing exponentially and requires careful consideration before implementation, and how brands are aligned with marketing strategies and based on the products of the companies.

After the introduction of Covid-19, the need for more concrete and directional information from the medical community and other healthcare system stakeholders has increased, because businesses are increasingly linking up with health care professionals.

Traditionally, healthcare providers or experts in the field can prescribe any number of treatments, assuming the drug would eventually be available for the patient, yet with technology advancements like BD, the healthcare stakeholders have to understand the manufacturing and supply chain process as well since treatment begins with enough clinical and scientific data.

4.3. Digital Transformation and Data Security

The feedback we received from the participants showed all companies are embracing digital transformations where digital data storage, processing, and exchange continue to become more challenging.

In addition to major reputational concerns and trust issues, these breaches may result in costs involved with mitigation and possibly increasingly harsh compliance and legal consequences.

Most companies today use various types of databases and other data stores for managing structured and unstructured data, depending on their business needs. The

EU's GDPR does not make a distinction between relational databases, data lakes, or file stores in its recent data protection regulations. Data governance and data protection have grown into a growing concern. When considering the BD V's, the most important one is volume when pharmaceutical companies have a mix of data types, including genetic data or any data set relevant to commercial operations.

Security threats of existing databases that can disrupt legitimate data access or lead to data loss, corruption, or sabotage are among the major concerns.

A special emphasis was also placed on evaluating whether administrators or accounts with excessive privileges inappropriately accessed sensitive data, as well as phishing, malware, or any other kinds of cyberattacks that might compromise legitimate data structures.

4.4. Business Objectives and Use Cases

Various use cases were discussed during the discussion, which shows that the main goal of using BD is not just to gather or generate enormous amounts of data, but to make sense of them and extract important information from them. Several companies are now actively utilizing BD with proper investments in data stewardship and data management to collect and manage data regarding customers, suppliers, transactions, and other relevant stakeholders in the marketplace such as competitors and supply chain players. BD was identified as the most pressing investment area along with the need to establish an effective and rewarding solution and to address the biggest concern associated with the management of a solid business development program.

In the most highlighted case studies, BD was associated with competitive intelligence analysis, NPS (Net Promoter Score), sales forecasts, and demand forecasts through market research initiatives.

In more than half of the companies, market research techniques were already in use to analyze specific information pieces and machine learning was employed to analyze public data and social media.

In addition, some examples were presented on how CRM can be analyzed based on insights and comments collected from event and interaction speech analytics, speech to text, and text analytics, thereby enabling more targeted, segmented, and comprehensive analysis. Furthermore, almost all participants shared that almost all companies still do not know how to analyze social media platforms to judge what customers are saying about them and what their competitors are saying.

Despite digitization, the Internet of things, and the proliferation of data sources, there are still many companies that are unable to maximize the value of and analyze the data using traditional data processing tools or traditional database management tools. In the case of large data sets, a variety of data formats can be used, including structured, semi-structured, and unstructured formats.

4.5. Method Discussions

Participants' responses and observations provided in the focus group interviews also brought to light a range of other important points to explain during this section of the report.

Below is a list of and explanation of the key topics to each of the questions presented, as well as an analysis of words that appear most frequently in the answers given.

A summary of all collected answers from focus group discussions is presented below and is arranged into four major columns: questions, themes, and topics most relevant to the discussions.

In addition to highlighting examples of BD used in the industry, the focus group interviews revealed the strengths and weaknesses of BD in commercial terms for different pharma companies (see table 3).

Table 3. Summary of all topics addressed in the focus group discussions with further details of the participants' comments and observations about all the topics discussed

Topic areas	Comments and observations
Data Strategy and Management	In terms of data management techniques and having the right internal skills, P&L and forecast analysis are key areas for big data analysis.
	The use of Big Data has been fundamental to the development of customer journeys for rare diseases, as well as new product launch scenarios when requesting support from commercial teams.
	Commercial analytics tasks related to wholesaler invoice reconciliation and wholesaler account management often involve large amounts of data integration.
	From a commercial excellence perspective, advanced analyses of customer reach, response curves, and key customers, as well as the distribution of field force by territory, are important in oncology companies. Here BD has its input, too.
	Various pre-processing operations are designed to address various defects in raw data collected so that a foundation for data analysis can be laid. There can be multiple sources involved in data, for example, so several formats are possible. In commercial operations or the commercial data management team, these activities are typically performed by data cleaning, data integration, and data transformation.
	It is usually not possible for data strategies to deal with different complexities, and there should be a clear understanding of the flow and potential difficulties of the deployment, maintenance, monitoring, and management processes.
Digital and Excellence	Pharma companies' futures are not only threatened by digital disruption, but it's their only hope, especially in a post-pandemic era. Businesses that accept that premise will be able to make rational and inevitable decisions that previously seemed nonacceptable.
	An attention-grabbing message and engaging engagement methods produced the best results from a customer-centric perspective, where key customers are most likely to benefit from centered actions.
	In contrast to harmonized, streamlined, and common processes for all countries, there are still different approaches and languages for the same commercial processes.
	Operational teams need to collect different data sets for omnichannel and cross-channel strategies that are evolving in a way that is aligned with communication trends and poses new challenges to the top management.
Systems and tools	In Commercial, tools play a critical role in the process's execution, and Big Data is a key component of operational and analytical effectiveness.
	The following processes are most relevant for commercial organizations at medium-sized pharmaceutical companies: customer continuous communication, forecasts, events management, sales incentives and rewards, content management, and analytical systems that can perform data mining and identify opportunities.
	A multichannel digital strategy relies on the integration of all sales channels and company data to determine marketing effectiveness and best practices. Accurate data at the right time is a critical component of a successful multichannel digital strategy.
Covid Impacts	Thousands of medical specimens and supplies travel between hospitals and clinics every day. COVID-19 vaccines, like other life-saving medicines, travel in transit, and keeping track of every item is imperative both to your health and the health of the healthcare system during this difficult time. It is important to communicate specific pieces of evidence and information pieces to our partners to be able to craft a successful commercial strategy. By harnessing these capabilities, the industry responds to global pandemics more effectively with data capture.
General considerations	When taking on a big data project, there must be a lot of adjustments, including new technology infrastructure, as well as a new way to capture, process, and use data.

	Table 5 (continued)
	There are increasing numbers of public and private organizations releasing publicly available datasets, and having teams equipped with the right knowledge to understand, process, and integrate data from various sources is essential.
	Most companies within the social network community have access to public and custom datasets through APIs that are generally designed to make data accessible.
Challenges and limitations	In addition to the volume issue, other challenges highlighted included the importance of real-time, predictive, and customer-centric analytics. In addition, you should consider defining business priorities by focusing on the most important activities that create value for the business.
	Especially for medium-sized or small pharma companies, securing a huge amount of continuously changing data can be extremely difficult, since their IT infrastructures cannot hold all of the data they need. It adds to the difficulties caused by big data being processed in real-time most of the time.
	There is no way to predict which platforms, applications, or methods will be best in the future when big data solutions proliferate. The need for big data solutions must be kept in mind by companies and the educational programs relating to commercial strategies, procurement, and even legal and compliance, should be available to different functions like IT and procurement.
Innovation	The goal of many companies is to extract value from huge sets of data. One way to do this is by using advanced analytics, which offers algorithms for performing complex analytics on either structured or unstructured data.
	We presented descriptive analytics as an important innovation point for several processes and projects, and it is typically used in promotional and health care professional engagement recommendation engines, as in CRM and Advanced analytics recommendation systems.
	An attendee at the meeting spoke of findings from a current pilot project that drove a series of innovative discussions and innovative analytics to discover that an unexplained variation in the evaluation of external stakeholders' weight involved research, sponsorship, donations, and grants, and engagement opportunities among other factors.
	Optimization analytics was another key innovation that assists with decision-making by answering various questions and is used by companies to optimize their inventory and supply chains, logistics, forecasts, and advertising strategies.
	Listening exercises provide insights into brands' social media participation strategies and plans as well as assess whether external stakeholders are actively participating in the process. BD participation has let us identify the best practices and continuously improve.
	Other examples in competitive advantage assessment projects included competitive intelligence, cyclical marketing campaigns, and Predictive analytics capabilities that could predict customer attributes, behaviors, or outcomes.

5. Conclusion

In this research study, we presented a brief study of innovation fundamentals and theories and contribute to the review of basic ideas related to the innovation process in the pharmaceutical industry, and we discussed the main issues for potentiating BD in the pharmaceutical industry.

The current BD applications in commercial pharmaceutical functions, trends in the future, and key characteristics of existing applications were therefore assessed. It was also possible to assess the connections and relationships which suggest BD and Digital drive innovation and can be considered as a valuable innovation source.

The study included 26 participants from various industries and business functions and examines the implications of these changes regardless of whether BD technology is used within businesses.

Hence, we analyzed the focus group findings closely and supported the research of BD as an innovation source for pharmaceutical processes and areas considering the current adoption and application of this technology.

BD will radically transform the pharmaceutical industry in the upcoming years, where enterprises that embrace this new technology to improve their working processes will have a bright future. Based on the focus group analysis, we were able

to understand that business development is fundamental to commercial organizations, especially when it comes to encouraging innovation and the implementation of new business practices.

In this context, it is important to manage two elements simultaneously: people, technology, and knowledge. Data management, skills development, information sharing, and recognition are core to promoting BD initiatives and projects because people are the source of knowledge.

It is also essential to implement mechanisms for systematic business development processes and generate insights that allow and empower employees' competencies and skills, as well as technological platforms that allow them to manage business development. In addition, it is necessary to emphasize the importance of BD as a tool that supports problem-solving and decision-making and identifies new approaches and routines required for changes in management practices and organizational structure that will lead to more competitive strategies and better performance.

In addition, all the interviews also led us to conclude that all participants in the study had a high amount of knowledge about the industry and business development, which allowed the research study to reach a good representation of the research questions.

Several pharmaceutical companies have launched BD programs with a wide range of services and processes in the last few years. Due to an increase in market challenges and competition pressure within the sector, most organizations have redefined their commercial business models and strategies.

As the demands for performance, agility, results, and value-driven outcomes grew from organizational leaders, the efficiency in sales and marketing operations had to reflect these changes, and here we can conclude that using our selected methodology, BD should be considered and implemented to drive innovation in pharmaceutical commercial modes.

Several processes and activities associated with commercial and digital applied to the pharmaceutical industry are common today. A new BD program requires organizational maturity, where core processes, such as the ones discussed in the focus group discussions, require financial and human investments.

References

1. Pesqueira A, Sousa MJ. (2020) Pharmaceuticals and Life Sciences: Role of Competitive Intelligence in Innovation. InHandbook of Research on Emerging Technologies for Effective Project Management, pp. 237-254. IGI Global, https://www.igi-global.com/chapter/pharmaceuticals-and-life-sciences/239222, last accessed 2021/06/26.
2. Ringberg T, Reihlen M, Rydén P. (2019) The technology-mindset interactions: Leading to incremental, radical or revolutionary innovations. 2019 May 1;79: pp.102-13. Industrial Marketing Management, https://www.sciencedirect.com/science/article/pii/S0019850118304218, last accessed 2021/05/24.
3. Laouisset DE. (2021) Algeria import substitution policy: the case of the pharmaceutical industry. les cahiers du cread. Apr 20;37(1):5-39, https://www.ajol.info/index.php/cread/article/view/206142/194380, last accessed 2021/07/14.
4. Lahane S, Kant R, Shankar R. (2020) Circular supply chain management: A state-of-art review and future opportunities. Journal of Cleaner Production.
5. Bibri SE. (2019) On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review. Dec;6(1): pp.1-64. Journal of Big Data, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0182-7, last accessed 2021/07/16.

6. Zhang D, Pee LG, Cui L. (2021) Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. 2021 Apr 1; pp. 57:102304. International Journal of Information Management.
7. Wiener M, Saunders C, Marabelli M. (2020) Big-data business models: A critical literature review and multiperspective research framework. Mar;35(1): pp.66-91. Journal of Information Technology, https://scholar.google.com/scholar?output=instlink&q=info:zxQHdZOuuKkJ:scholar.google.com/&hl=en&as_sdt=0,5&scillfp=11377637586559752519&oi=lle, last accessed 2021/06/12.
8. Verma S. (2017) Big Data and advanced analytics: Architecture, techniques, applications, and challenges. 2017 Oct 1;4(4): pp.21-47. International Journal of Business Analytics (IJBAN).
9. Mariani MM, Nambisan S. (2021) Innovation analytics and digital innovation experimentation: the rise of research-driven online review platforms. Nov 1; pp. 172:121009. Technological Forecasting and Social Change.
10. Preim B, Lawonn K. (2020) A survey of visual analytics for public health. Feb Vol. 39, No. 1, pp. 543-580. InComputer Graphics Forum, https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/cgf.13891, last accessed 2021/06/13.
11. Büyüközkan G, Göçer F. (2018) Digital Supply Chain: Literature review and a proposed framework for future research. Computers in Industry. May 1;97: pp. 157-77.
12. Zhao J, Xue F, Khan S, Khatib SF. (2021) Consumer behavior analysis for business development. Aggression and Violent Behavior. Vol. 1 pp.101591.
13. Ciruela-Lorenzo AM, Del-Aguila-Obra AR, Padilla-Meléndez A, Plaza-Angulo JJ. (2020) Digitalization of agri-cooperatives in the smart agriculture context. proposal of a digital diagnosis tool. Sustainability. 12(4): pp.1325, https://www.mdpi.com/2071-1050/12/4/1325/pdf, last accessed 2021/08/17.
14. Saqib N. (2020) Positioning–a literature review. Jul 11. PSU Research Review. https://www.emerald.com/insight/content/doi/10.1108/PRR-06-2019-0016/full/html
15. Vickers I, Lyon F, Sepulveda L, McMullin C. (2017) Public service innovation and multiple institutional logics: The case of hybrid social enterprise providers of health and wellbeing. Dec 1;46(10): pp.1755-68. Research Policy, https://www.sciencedirect.com/science/article/pii/S0048733317301373, last accessed 2021/07/10.
16. Aghmiuni SK, Siyal S, Wang Q, Duan Y. (2020) Assessment of factors affecting innovation policy in biotechnology. Jul 1;5(3): pp. 180-90. Journal of Innovation & Knowledge. https://www.sciencedirect.com/science/article/pii/S2444569X1930054X, last accessed 2021/08/26.
17. Teece DJ. (2018) Dynamic capabilities as (workable) management systems theory. 2018 May;24(3): pp.359-68. Journal of Management & Organization.
18. Orlandi LB, Zardini A, Rossignoli C. (2020) Organizational technological opportunism and social media: The deployment of social media analytics to sense and respond to technological discontinuities. May 1;112:385-95. Journal of Business Research.
19. Lüdeke-Freund F. (2020) Sustainable entrepreneurship, innovation, and business models: Integrative framework and propositions for future research. Feb;29(2): pp.665-81. Business Strategy and the Environment, https://onlinelibrary.wiley.com/doi/pdf/10.1002/bse.2396, last accessed 2021/07/26.
20. Pratono AH, Darmasetiawan NK, Yudiarso A, Jeong BG. (2019) Achieving sustainable competitive advantage through green entrepreneurial orientation and market orientation: The role of inter-organizational learning. Mar 11. The Bottom Line, http://repository.ubaya.ac.id/34237/1/AS71694623652659215479447119 31_content_1.pdf, last accessed 2021/07/16.
21. Farzaneh M, Ghasemzadeh P, Nazari JA, Mehralian G. (2020) Contributory role of dynamic capabilities in the relationship between organizational learning and innovation performance. 2020 Jun 3. European Journal of Innovation Management.
22. Chege SM, Wang D. (2020) The impact of entrepreneurs' environmental analysis strategy on organizational performance. Jul 1;77: pp.113-25. Journal of Rural Studies.
23. Lee C, Hallak R, Sardeshmukh SR. (2019) Creativity and innovation in the restaurant sector: Supply-side processes and barriers to implementation. Jul 1;31:pp. 54-62. Tourism Management Perspectives.
24. Panda DK. (2017) Coevolution and coexistence of cooperation and competition in inter-organizational collaboration: Evidence from Indian management consulting industry. Feb 20. Journal of Global Operations and Strategic Sourcing.

25. Ross JW, Beath CM, Mocker M. (2019) Designed for digital: How to architect your business for sustained success.2019 Sep 24. MIT Press.
26. Dubey R, Gunasekaran A, Childe SJ, Blome C, Papadopoulos T. (2019) Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view, and big data culture. Apr;30(2): pp. 341-61. British Journal of Management, https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/13321/Final_BJM_Revised%20Version_Authors%20Affiliations.pdf?sequence=1, last accessed 2021/07/26.
27. Mikalef P, Boura M, Lekakos G, Krogstie J. (2019) Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. 2019 Apr;30(2): pp. 272-98. British Journal of Management, https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2631308/Big%2BData%2BAnalytics%2BCapabilities%2BR2.pdf?sequence=1, last accessed 2021/08/12.
28. Faisal-E-Alam M. (2020) The Influence of Quality on Consumers' Purchase Intention between Local and Multinational Cosmetic Firm. 3(1): pp.01-11. Journal of International Business and Management. https://www.researchgate.net/profile/Md-Faisal-E-Alam/publication/338498835_The_Influence_of_Quality_on_Consumers'_Purchase_Intention_between_Local_and_Multinational_Cosmetic_Firm/links/5e26f95992851c89c9b5c289/The-Influence-of-Quality-on-Consumers-Purchase-Intention-between-Local-and-Multinational-Cosmetic-Firm.pdf, last accessed 2021/07/16.
29. Rey JE.(2020) Career-Advising from the Primary Role Academic Advisor's Viewpoint: A Qualitative Case Study. Rowan University. https://rdw.rowan.edu/cgi/viewcontent.cgi?article=3840&context=etd, last accessed 2021/07/15.

Preliminary performance evaluation and verification of digital terrestrial television signal propagation

Leboli Zachia Thamae[1][0000-0002-2886-6549], Itumeleng J. Potsanyane, Mpho W. Mokhetsengoane

[1] Department of Physics & Electronics, National University of Lesotho, Roma 180, Lesotho

https://doi.org/10.33847/2686-8296.3.2_4

Received 20.10.2021/Revised 06.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. This article presents the computer simulation and field test measurement results on Channel 29 for the preliminary performance evaluation and verification of the newly-installed Lesotho digital terrestrial television network based on DVB-T2 standard following the guidelines and techniques specified by the ITU-R BT.2035-2. It evaluates, at predetermined outdoor locations for fixed and mobile reception, parameters such as received signal strength, signal quality, bit-error rate (BER) and threshold-of-visibility (ToV) together with TV signal decoding (observation of screen artefacts) for quasi error-free reception. The results indicate that at over 97% of the test sites/points at the university town of Roma, the main Berea Plateau transmitter from the capital city (Maseru) broadcasts digital television service with enough signal level and quality to be properly decoded. The measured signal strength threshold ranges above -50 dBm for good reception, -64 dBm to -50 dBm for acceptable reception and -69 dBm to -64 dBm for poor reception. With the noise floor at about -73 dBm, the minimum required C/N of around 23 dB for good reception and about 4 dB for ToV have been recorded. The relative values of minimum required respective signal strength and signal quality for ToV obtained from the set-top box are 33% and 18% for stationary reception, while they give 37% and 20% for mobile reception.

Keywords: DTT performance, mobile reception, signal strength, stationary reception, threshold of visibility.

1. Introduction

The television (TV) broadcasting industry in Lesotho has been dominated by the giant South African-based MultiChoice with its digital satellite service under the DSTv brand. The local public service broadcaster, Lesotho National Broadcasting Services (LNBS), operates a single analogue terrestrial TV channel that broadcasts live for only 7 hours per day, and it is also replicated on DSTv Channel 292. From 2006, with the agreed mandate from the International Telecommunications Union (ITU) for countries in Region 1 to migrate to digital terrestrial TV (DTT) broadcasting by 17 June 2015, efforts were put in motion for the country to be ready for analogue switch-off by the international deadline, though with little success due to various migration challenges [1]. However, since then, there has been demonstrable progress, especially in the building of transmission infrastructure based on the second-generation Digital Video Broadcasting – Terrestrial (DVB-T2) standard at different locations around the country as indicated on Fig. 1. Most of the transmitter locations were already in use for terrestrial analog FM radio broadcasting and are able to communicate via microwave links.

As evidenced by the topographical map of Fig. 1, about 70% of the country's

terrain is characterized by rugged mountains and elevations in excess of 1400 m, presenting a challenge for optimal location of broadcasting transmitters. The LNBS's optimized DTT network configuration comprises 20 transmitter sites to cover the country's approximately 30,355 km2 area, which is completely surrounded by South Africa. The main transmitter station for coverage provision in the capital city Maseru is called 'Berea Plateau', located in the western lowlands at an elevation of 1829 m. At the time of writing, only the main Berea Plateau transmitter station was operating in Channel 29 (538 MHz) under test conditions with set-top boxes (STBs) from Reutech (South Africa) and StarTimes (China).

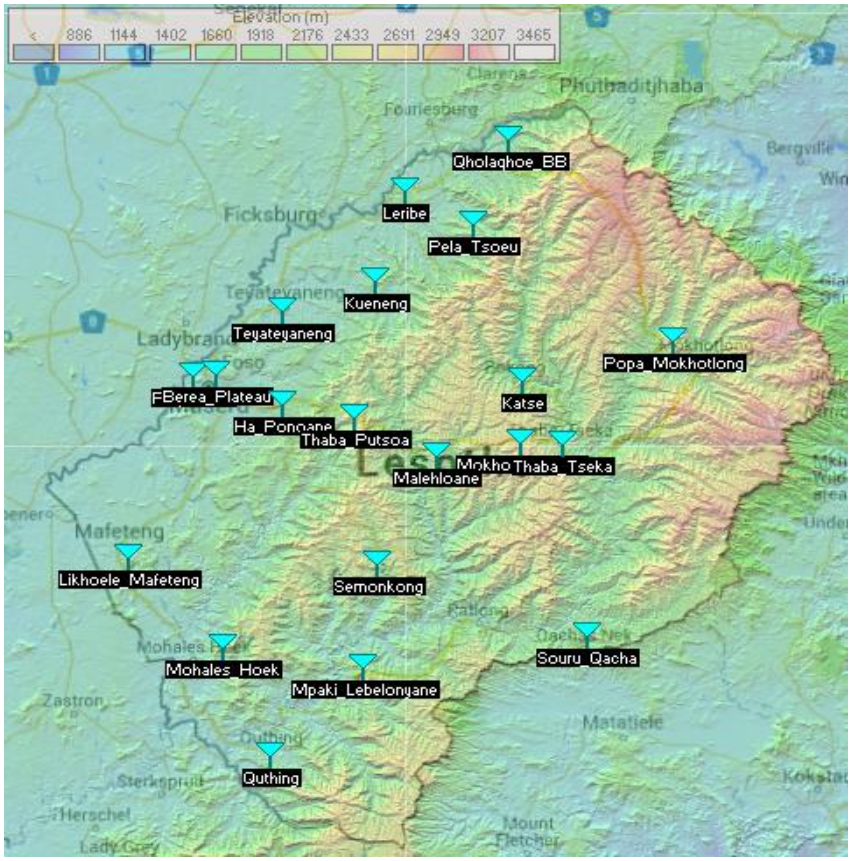


Fig. 9. Topographical map of Lesotho showing locations of newly-built DTT transmitter stations

The LNBS DTT network topology is shown in Figure 2, adapted on the French Eutelsat geo-stationary satellite link as the primary distribution network and the 20 terrestrial transmitter stations forming the secondary distribution network, configured as a multiple frequency network (MFN) consisting of individual transmitters with each transmitter on a different frequency. The head-end system and uplink satellite dish are located at the main transmitter site at Berea Plateau. The DTT signal is converted to DVB-S2 and encrypted for uplink and the resulting transport stream is fed into the primary distribution network by means of a microwave satellite feed at 8° West for easy implementation of its point to multipoint distribution [2]. The downlink DTT signal from the satellite is then decrypted and redistributed as DVB-T2 on a free-to-air (FTA) basis by the terrestrial transmitter stations scattered around the country as in Figure 1 to cover various service areas.

Experience from early adopters has shown that first-time operation of new DTT broadcasting network, with limited time for planning and extensive testing due to accelerated digital switchover, can be technically challenging [3]. This can lead to reception problems at shadowed areas without coverage, known as black spots, due to an uneven terrain and non-optimal tower erection. Hence this study aims to measure actual service versus predicted coverage by modelling the DTT network for Lesotho and performing the first verification of area coverage for correct decoding of digital TV signal against the basic parameters stated in the ITU-R BT.2035-2 guidelines and techniques [4]. It employs useful comparison of computer simulated predictions using Radio Mobile software [5-7] with practical field strength measurements in validating network performance. This is followed by the determination of minimum signal strength and carrier-to-noise ratio (C/N) required to attain quasi-error free (QEF) reception under real propagation conditions. As an indication of proper digital signal decoding, coverage evaluation with the observation of the image on the TV screen is performed in predetermined sites.

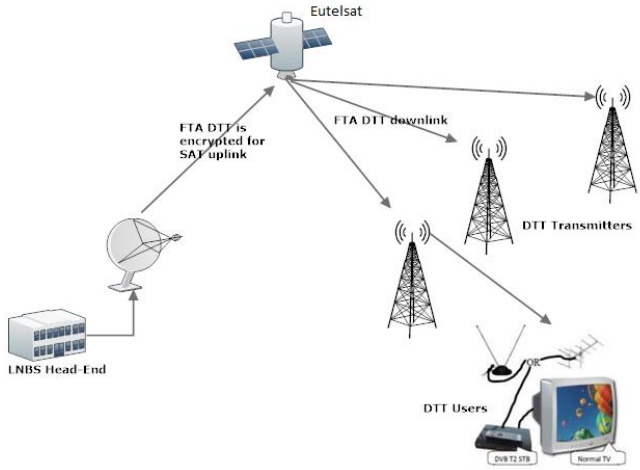


Fig. 10**.** LNBS DTT network topology.

The results of the study, if replicated throughout the country, will help identify problem sites within the network coverage area or its peripheries where DTT signal is poorly available, and act as aid to confirm DTT broadcasting coverage requirements, and assist in future planning to maximize coverage and service quality. The accurate knowledge of the quality of DTT coverage to be provided will be key to the implementation of digital television services and give credibility to the service provider before total analogue switch-off.

2. Review of DTT Network Simulations and Measurements

The transmitter network structure for coverage planning involves optimizing the locations of transmitters and gap-fillers in regions where the service is poor. Two reception conditions, fixed (or stationary) and portable (or mobile) are usually considered in DTT broadcasting [11]. Receiver antenna heights range around 10 m and 3 m for fixed reception and 1.5 m for portable reception with directional and omni-directional radiation patterns, respectively [10, 13]. As a rule of thumb, if the required C/N is attained 99% of time, the location is deemed to be covered [11]. Furthermore, coverage of a small area, typically 100 m x 100 m, is classified as "good" if at least 95% of receiving locations within it are covered, or "acceptable" if at least 70% of locations within it are covered [14].

Due to continuing co-existence with analogue services in many regions which failed to meet the ITU deadline, tighter margins on the signal strength and interference are placed on digital broadcasting services and thus, they require higher prediction accuracy than analogue networks [15]. Moreover, picture quality deteriorates sharply over a very small range of signal levels and thus digital television service coverage is characterized by a very rapid transition from near-perfect reception to no-reception at all [14]. As such, various radio-frequency modeling and simulation software packages are available for use in predicting and analyzing the global coverage probability within a target service area. In this study, the Radio Mobile [5, 7] freeware radio propagation simulation program is utilized to model the Lesotho DTT network and allow analytical study of simulated reception conditions to be made.

The models for propagation prediction can be classified as semi-empirical and fully deterministic with varying requirements for initial set-up and provided accuracy [15]. Radio Mobile, which can be classified as deterministic, is based on the Longley Rice Irregular Terrain propagation model and uses high-resolution topographical data from the Shuttle Radar Topography Mission (SRTM) elevation maps to take into account the actual terrain of the area being simulated [6, 7]. It enables merging of elevation contours and roads with maps and allows specification of crucial network information such as location of transmitter stations (units), units' frequencies, transmit power, height, antenna gain, receiver sensitivity and feeder losses [5]. Path profiles with Fresnel zones and signal parameters for radio links between units can be examined, together with signal coverage patterns from individual or multiple transmitter units showing predicted receive signal levels.

Following computer simulation predictions, field test measurements and data collection (coverage, service, modes and channel characteristics) are usually undertaken with the view to ascertain the actual service and be able to improve system performance [4]. Due to local clutter effects (buildings, trees and other obstructions), the reception environment for practical measurements is usually classified in terms of urban, suburban or rural, depending on various characteristics such as densely-populated buildings with high-rise structures, disperse buildings or sparsely-populated structures [4]. Various test parameters for quasi-error free reception or correct TV signal decoding for digital terrestrial television have been measured in several places. The test parameters (and their typical measured values in suburban areas) include signal spectrum, signal or carrier power (dBm), signal field strength (above 50 dBμV/m for stationary reception and above 73 dBμV/m for portable reception), receiver equivalent noise floor (-73 dBm), carrier-to-noise ratio (C/N > 20 dB), bit error rate (BER < 2x10-4 after Viterbi decoder), latitude/longitude and TV set image [10, 12, 14, 16, 17].

3. Simulated LNBS Network Performance

The LNBS's DTT network is simulated using Radio Mobile by first producing the elevation map of Lesotho centred at 29°37'24.0"S, 28°12'16.0"E with 270 km height and 514 x 514 pixels to cover the entire country. An Internet GoogleMap roadmap is merged with the elevation map and the 20 transmitter units are defined to produce the map of Figure 1. The main transmitter station at Berea Plateau, the gap-filler at Ponoane (Roma) and a modeled portable unit with the parameters shown in Table 1 are used for simulation and illustration purposes.

A. Simulated Path Profiles

The Radio Link toolbar is used to examine the simulated path ground profiles and link budget analysis between the main transmitter and the modelled portable unit placed 22 km away at the main entrance to the National University of Lesotho (NUL) Roma campus as demonstrated by Figure 3. The green colour on ground path

indicates that the signal level at the receiver is above the minimum margin required for reception. These settings are made relative to the receiver sensitivity, with an intermediate band of signal levels being shown as yellow for ±3 dB, >3 dB as green, and <-3 dB showing red [6]. The figure further demonstrates information regarding the terrain elevation difference, minimum clearance, location of the obstruction giving the worst Fresnel clearance, distance between the units, propagation mode, total propagation loss, field strength, signal level at the receiver and attained link margin.

Table 1. The operating parameters used to define main Transmitter, Gap-Filter and Portable Units

Parameters	Berea Plateau (Main Unit-Maseru)	Ponoane/Roma (Gap-filler)	Portable Unit
Location	29°19'50.8"S, 27°32'25.2"E	29°25'07.3"S, 27°44'35.9"E	29°27'05.0"S, 27°43'13.5"E
Frequency (MHz)	470 – 862	470 – 862	470 – 862
Channel Frequency	29 (538 MHz)	45 (666 MHz)	-
Transmit Power (W)	5000	600	1
Antenna	Omni	Omni	Omni
Antenna Gain (dB)	12.3	7	2
Antenna Height (m) agl	200	30	1.8
Receiver threshold (μV)	1.0	0.5	0.5
Line loss (dB)	1.0	0.5	0

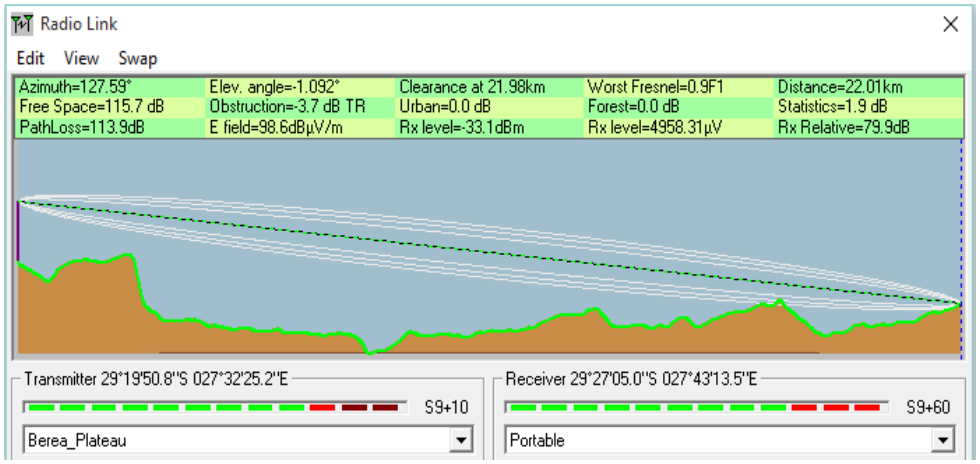


Fig. 11. Simulated path profile between the main transmitter unit and the portable unit.

B. Simulated Coverage Plots

The simulated transmitter coverage plots from Radio Mobile software can be exported to Google Earth to explore the predicted coverage in a three-dimensional (3D) environment. Coverage plots from individual transmitter stations around the country can be aggregated to predict the national DTT signal coverage. For instance, the simulated national coverage map combining all 20 transmitters installed by LNBS is given in Fig. 4. It can be seen that signal coverage is more uniform and strong (red colour) in the western densely-populated lowlands (Butha-Buthe to Maseru to Mohale's Hoek) than in the central and eastern sparsely-populated highlands.

A zoomed-in view of the signal coverage plot for the university town of Roma valley is shown in Fig. 5, depicting TV signal strength distribution from the main Berea Plateau transmitter station. The highlighted spots with circles indicate areas where the signal strength is relatively low, leading to potential coverage gaps with poor signal reception from the main transmitter, especially around the steep hills and the nearby Liphiring river where the valley is deeper. However, a gap-filler transmitter station at the nearby

Ponoane Hill has been installed to provide signal coverage to these spots that are not properly covered by the main transmitter as illustrated in Fig. 6.

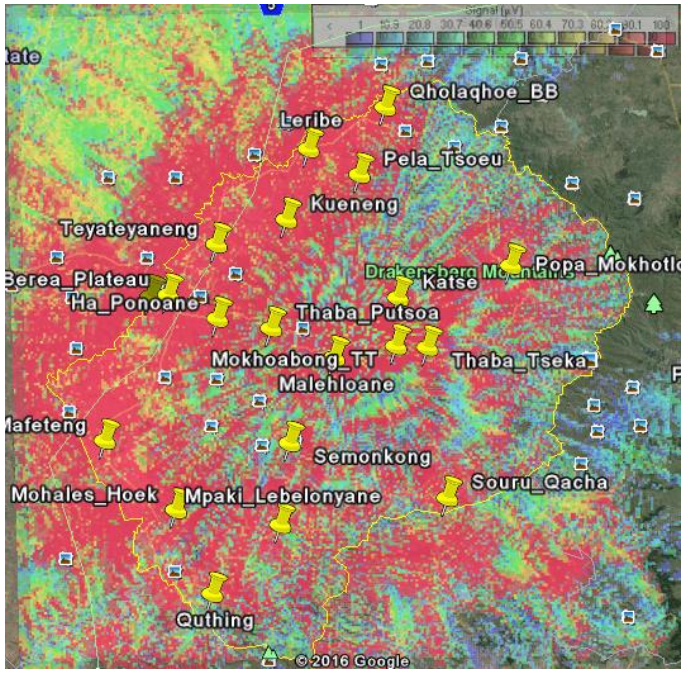


Fig. 4. Simulated Lesotho DTT signal coverage map.

Fig. 5. Simulated coverage plot of the university town of Roma from the Berea Plateau transmitter.

C. Simulated Route Reception

The Radio Mobile software is also able to simulate portable signal reception for demonstration of mobile TV reception. For illustrative purposes, a route from Maseru

via Thetsane Industrial Area to Roma is simulated as shown in Fig. 7 (a). With Berea Plateau as the transmitter, Fig. 7 (a) shows that the simulated DTT signal reception varies along the route as indicated by the highlighted areas, becoming smaller (red) around Ha Tsolo and the Roma valley areas. An enlarged view of the simulated poor mobile TV signal reception (below -70 dBm) from Berea Plateau around the first spot at Ha Tsolo is shown in Fig. 7 (b), caused by the nearby obstructing Qoaling Hill towards the right. However, there are gap-fillers at Ratjomose and Ponoane to solve the respective coverage problems of Ha Tsolo and Roma with mobile reception. The improved reception at Ha Tsolo (around -25 dBm) from the Ratjomose gap-filler is illustrated in Fig. 8.

Fig. 6. Simulated coverage plot of the Roma valley from the Ponoane gap-filler transmitter station

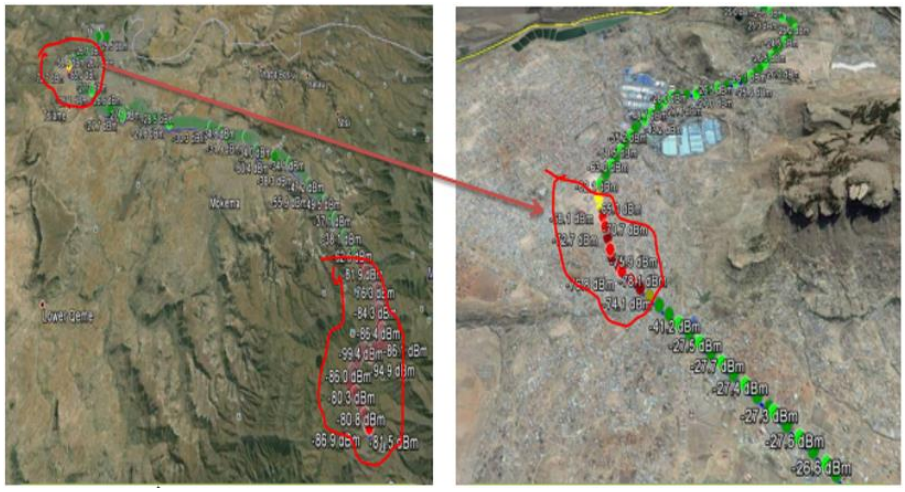


Fig. 7. (a) Simulated mobile route reception from Maseru to Roma (left); (b) The enlarged simulated poor DTT signal reception at Ha Tsolo from the main transmitter (right)

Fig. 8. Simulated improved DTT mobile signal reception at Ha Tsolo from the Ratjomose gap-filler

4. Signal Measurements and Analytical Evaluation

Field data acquisition to evaluate the network performance was carried out with a mobile measurement system set-up as indicated in Fig. 9, following the ITU-R BT.2035-2 guidelines and techniques [4]. The actual measurement system, which is mounted on a vehicle, consisted of a retractable antenna mast, 8-element grid mesh TV antenna, splitter, StarTimes Gemini+ set-top box, RIGOL DSA1030A spectrum analyzer, TV set, GPS receiver and laptop. The power for the equipment was provided by a 350W inverter connected to the van's 12-V cigarette lighter. The antenna mast is extendable to lift the antenna to 3m above ground level for stationary reception or 1.5m for mobile reception.

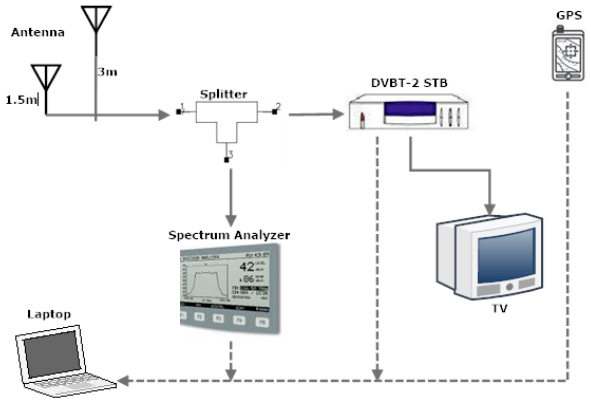
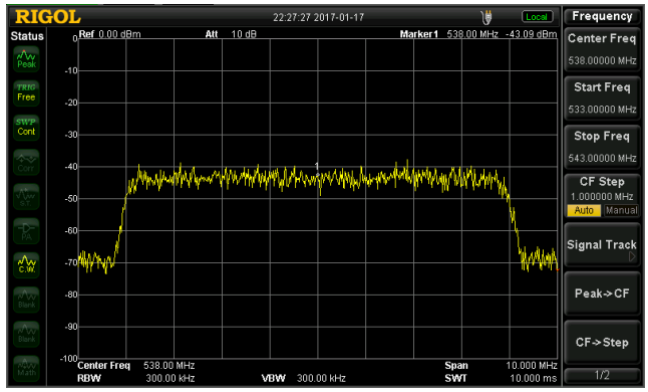
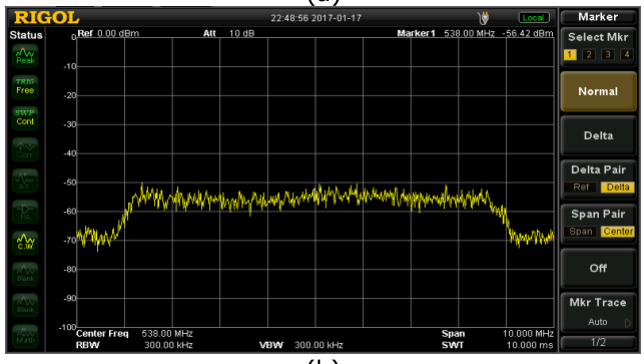


Fig. 9. Measurement setup for DTT signal capturing and analysis.

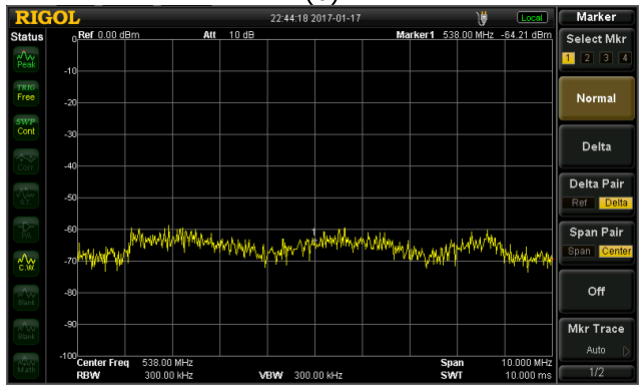
The spectrum analyzer captures and stores details of the signal spectrum, field strength and power received. Observation of the video on the TV screen from the connected STB is used for visual monitoring with the signal classified as either good (near perfect picture) or poor (no picture). The STB further provides measurements of signal strength and signal quality in percentages and BER that are displayable on the TV screen. The minimum values of these parameters (signal strength, signal quality and BER) needed for quasi-error free reception of signal are termed the threshold-of-visibility (ToV).



(a)



(b)



(c)

Fig. 10. Recorded narrow-band (10 MHz) displays of DTT signal spectrum at the selected test points for: (a) good reception, (b) acceptable reception and (c) poor reception

A. Measured DTTB Signal Analysis

The measured signal analysis was undertaken using the obtained signal strength or level, BER, and determination of C/N ratio. At various test points, the signal level and quality was high enough (> -53 dBm) for good quality video reception. At some test points, the signal level was lower than -64 dBm and not good enough for proper decoding of the video. Figure 10 illustrates the narrow-band displays captured from the spectrum analyzer for good, acceptable and poor reception based on the observation of the image on the TV screen and benchmarked with operating parameters in Table II. Another important system parameter measured is the minimum C/N required for QEF reception. At each test point, this parameter is determined by varying the attenuation on the received DTT signal until its strength reaches (ToV), i.e. it is low enough to be considered equal to the set-top box decoder noise floor (-73 dBm) and the picture is lost on the TV/laptop screen. The required minimum C/N averaged around 23 dB.

B. Stationary Reception Results

Two sample areas have been chosen for investigating coverage probabilities from the main transmitter at Berea Plateau relative to the simulated coverage plot of the Roma valley shown in Figure 5. The first area has 26 test points at 100 m intervals within the university campus (see Figure 11) and the second one has 35 tests points at 500 m intervals for villages around the university campus (see Figure 13). The test points have been selected to sample areas with good and poor coverage as predicted by the simulation of Figure 5. The test points are also classified as suburban with disperse buildings where the receiving antenna is mounted at roof top level (3 m). It should be noted that only the main Berea Plateau transmitter was working under test conditions during the study period.

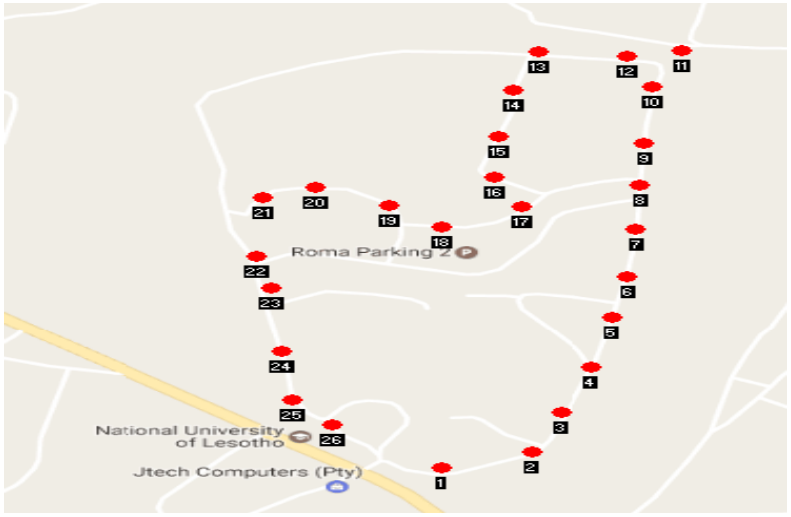


Fig. 11. DTT signal reception test points within university campus.

The measured signal properties for stationary reception test points within the university campus at Roma and for villages around the campus are illustrated in Figure 12 and Figure 14, respectively. The signal strength (%) and signal quality (%) have been captured from the STB readings while the signal power (dBm) was recorded by the spectrum analyzer. For the stationary reception results, the ToV was maintained when the signal strength was 33% or higher and the signal quality was 18% or higher

from the STB, with the signal power greater than -70 dBm. Hence for the 61 test points recorded, only 2 points (3%) reached 18% signal quality with – 70 dBm signal power, resulting in poor reception.

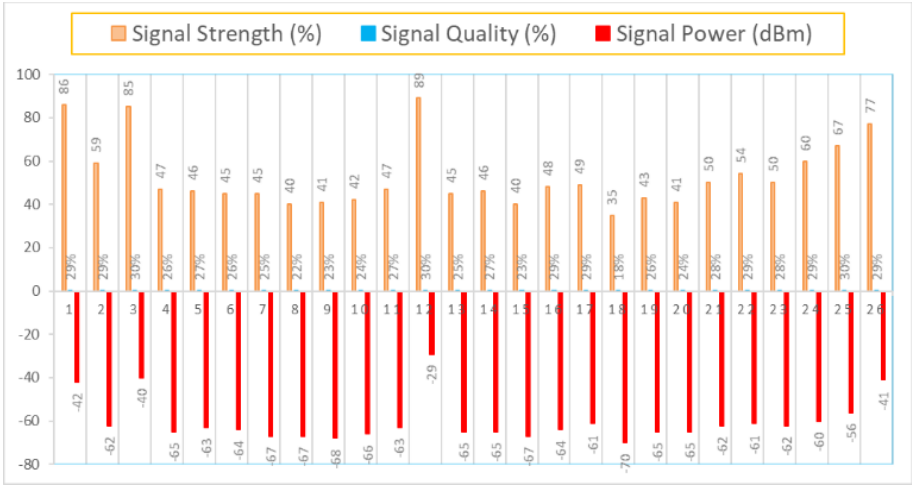


Fig. 12. Measured signal properties at test points within university campus.

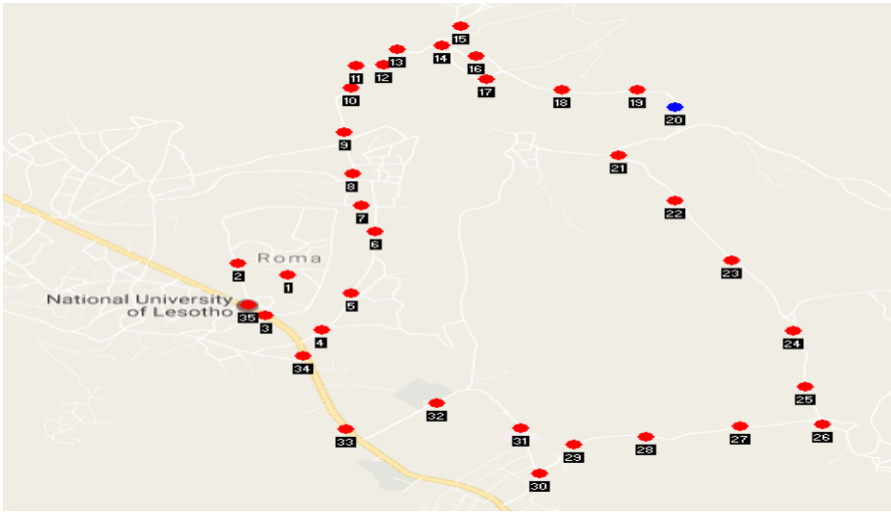


Fig. 13. DTT signal reception test points for villages east of the university campus.

Table 2. The operating parameters used for measurements

Reception Classification	STB Signal Strength (%)	STB Signal Quality (%)	STB BER	Spectrum Analyzer Signal Strength (dBm)
Good	> 50%	24% - 30%	~ $1x10^{-9}$	> -50 dBm
Acceptable	37% to 50%	14% - 24%	< $1x10^{-4}$	-50 dBm to -64 dBm
Poor	< 37%	< 14%	> $2x10^{-4}$	< -64 dBm to -69 dBm

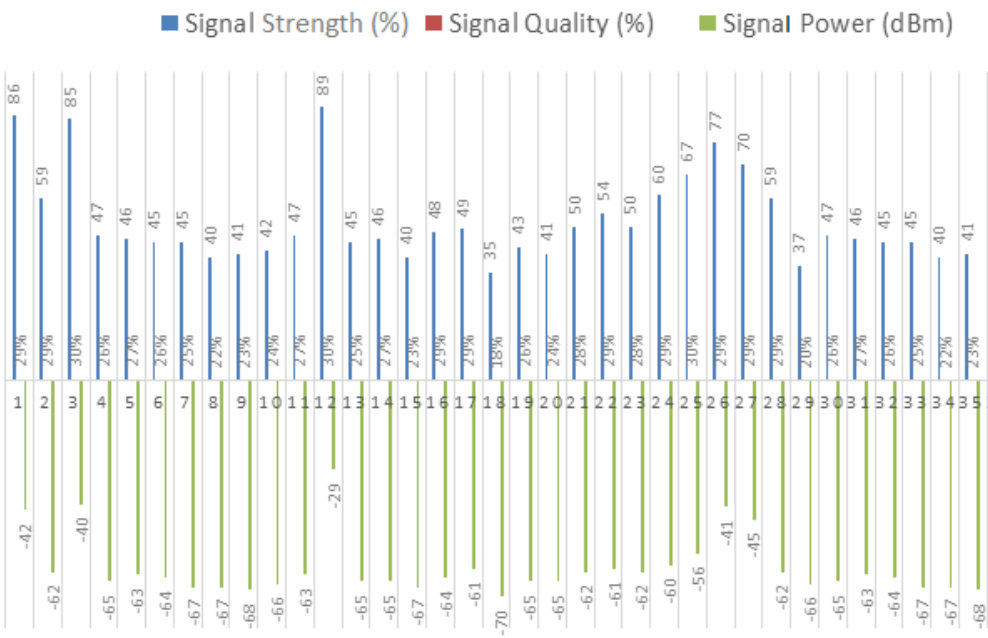


Fig. 14. Measured signal properties for test points in villages east of the university campus

Fig. 15. DTT mobile signal reception test points for the St. Michal's through Roma to Nyakosoba route

C. Mobile Reception Results

Measured signal properties for mobile or portable reception along the A5 route from St. Michaels to Nyokosoba simulated in Figure 15 were taken with the receiving antenna at 1.5 m above ground level. A total of 31 test points spaced at 500 m intervals along the route as illustrated were sampled. For the mobile reception results, the ToV was maintained when the signal strength was 37% or higher and the signal quality was 20% or higher from the STB. The results in Figure 16 indicate that about 6 points (numbered 21, 27 – 31) attained signal strength of less than 37% and signal quality of less than 20%. In fact, for test points 27 to 31, there was no signal recorded at all as the area towards Nyakosoba is in a valley shadowed by mountains. Thus, only about 81% of the test points for mobile route reception received good enough signal level and video quality.

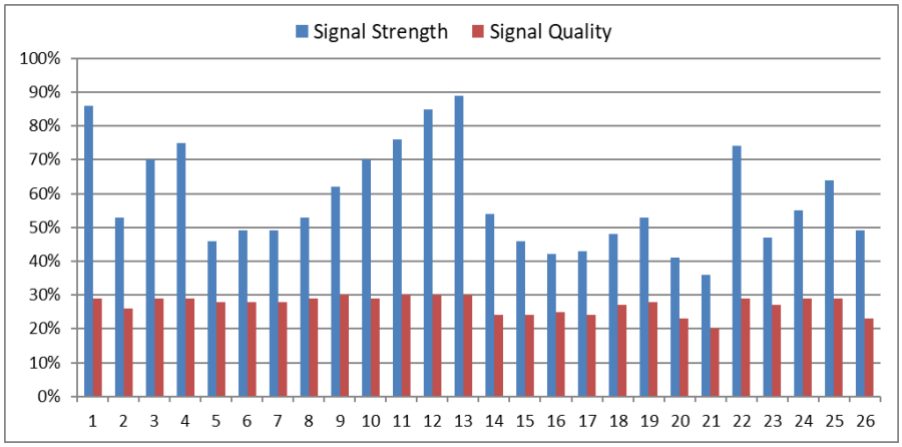


Fig. 16. Measured signal properties for test points in mobile reception along St Michael's to Nyakosoba

5. Conclusion

This study presented the first measurement and verification or validation campaign for Lesotho's digital terrestrial TV network signal reception and decoding, based on DVB-T2 and MPEG-4. The field testing objectives were to measure the actual service versus predicted coverage from computer simulations, and to collect data that will be useful in improving the Lesotho's DTT broadcasting system performance. The study found that at most tests points in the university town of Roma and nearby villages, which are within the coverage area of the main transmitter at Berea Plateau, the transmitter broadcasts the DTT signal with enough level and quality to be decoded without errors and properly viewed on TV screens. Values of the required minimum C/N for quasi-error free reception have been determined to average 23 dB for fixed reception. The study further confirmed the existence of coverage gaps or black spots within the selected measurement test points which are shadowed from the main transmitter by nearby hills and mountains and therefore have poor signal reception as demonstrated by the Radio Mobile simulation results. The reception problem in these shadow areas will be addressed by the gap-fillers once they become operational.

In order to obtain statistically significant results there must be enough data sample points measured to reflect the actual performance of the measured system. Practical considerations lead to a range of 30 to 100 sites, although reasonable statistical confidence intervals may require significantly more tests points, which will be done in future once the whole DTT network is operational.

References

1. L. Z. Thamae, "A review of Lesotho's digital migration challenges: Policy lessons from global and regional experiences," International Journal of Digital Television, vol. 6, pp. 331-346, 2015.
2. G. A. Jones, J. M. Defilippis, H. Hoffmann, and E. A. Williams, "Digital television station and network implementation," Proceedings of the IEEE, vol. 94, pp. 22-36, 2006.
3. N. H. Gunze, "Digital migration experience, challenges and success of Tanzania," in DTT Consultative Stakeholder Conference Maseru, Lesotho, 2014.
4. ITU, "Guidelines and techniques for the evaluation of digital terrestrial television broadcasting systems including assessment of their coverage areas," REPORT ITU-R BT.2035-2 2008.
5. I. D. Brown, "An introduction to Radio Mobile," RadCom, 2006.
6. I. D. Brown, "Radio Mobile - What can it do for you?" AntenneX, 2009.
7. B. J. Henderson, "Radio Mobile program operating guide," VE6ZS 2011.
8. J. M. Fernández, J. Capdevila, R. García, S. Cabanillas, S. Mata, and A. Mansilla, "Single frequency network for digital video broadcasting," RETEVISION S. A., Spain 2000.
9. M. Venter, "Network planning considerations and rollout of DTT," in Lesotho DTT Consultative Stakeholders Conference Maseru, Lesotho, 2014.
10. U. Ladebusch and C. A. Liss, "Terrestrial DVB (DVB-T): A broadcast technology for stationary portable and mobile use," Proceedings of the IEEE, vol. 94, pp. 183-193, 2006.
11. C. Werk, "Coverage aspects of digital terrestrial television broadcasting," EBU Technical Review, 1996.
12. ABA, Digital Terrestrial Television Broadcasting Planning Handbook. Canberra: Australian Broadcasting Authority, 2005.
13. F. S. Caluyo and J. C. Dela-Cruz, "Antenna characterization and determination of path loss exponents for 677 MHz channel using fixed and portable digital terrestrial television," Progress In Electromagnetics Research C, vol. 29, pp. 149-161 2012.
14. EBU, "Terrestrial Digital Television Planning and Implementation Considerations," Geneva 2013.
15. B. Belloul and S. Saunders, "Accurate coverage prediction and optimization - for digital broadcasting," EBU Technical Review - Spectrum Planning, April 2014.
16. A. Arrinda, J. L. Ordiales, M. M. Velez, and P. Angueira, "Digital terrestrial television (COFDM 8K System) field trials and coverage measurements in Spain," University of the Basque Country, 1999.
17. J. M. Fernández, J. Capdevila, R. García, S. Cabanillas, S. Mata, and A. Mansilla, "Single frequency network for digital video broadcasting," RETEVISION S. A., Spain 1999.

Determinants, Barriers and Strategies of Digital Transformation Adoption in a Developing Country Covid-19 era

Kingsley Ofosu-Ampong[0000-0003-0561-6376]

Baldwin University College, P. O. Box 19872, Accra – North, Ghana

https://doi.org/10.33847/2686-8296.3.2_5

Received 09.10.2021/Revised 03.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. The purpose of this paper is to examine the determinants and strategies of digital transformation adoption (DTA) in a developing country context through the lens of price value, hedonic motivation, inherent innovativeness and technology readiness. The study also investigates the impact of COVID-19 on banks in Ghana from the managers' perspectives and provide possible solutions for banks' successful transitioning and uptake of digital transformation in a post COVID-19 era. The study was carried out using a mixed-method approach from banks in Ghana. The findings of the study revealed that customers' technology acceptance and adoption of innovation is fraught with challenges. At the same time, employees/banks struggled to adjust to new technologies during the COVID-19 pandemic. Further, the findings indicate that price value, inherent innovativeness and technology readiness were the significant factors in DTA. Conversely, hedonic motivation was an insignificant factor in a developing country context. The paper concludes with a conceptual model for emergency digital transformation to respond to future pandemics.

Keywords: Digital transformation, Determinants, Manager's perspective, Covid-19 banking, Developing countries, Strategies.

1. Introduction

The future is digital. As such digital technology is globally recognised as a reliable tool for enabling bank transformation and development and interacting with customers. In recent times, digital transformation has been widely observed in operations management, information technology, and business for organisation and strategic changes. It has been described as a new development in information systems (IS) research [1]. Digital transformation in this context refers to leveraging digital technologies to (re)define the value position of an organisation, and it involves a new organisational identity [2].

Extant literature on digital transformation in developing countries is scanty and remain limited. However, the factors of digital transformation in the developed world are well-grounded in IS literature. To a large extent, the digital transformation happening in developed countries differed from developing countries. For instance, the technological infrastructures acquired by banks [1] in developed countries such as US or UK is more sophisticated and advanced than banks in a developing country like Ghana. Further, as widely noted in IS literature, research findings from developing countries do not directly apply to developed countries and vice versa due to socio-cultural factors [3]. Also, since the COVID-19 pandemic was declared a Public Health Emergency of International Concern by the World Health Organization on 30 January 2020, both developed and developing countries have embraced technology and dynamic capabilities as an enabler of agility and resilience and adopted different COVID-19 protocols to stay economically active and gain high profits. However, the factors associated with developing countries digital technologies in a banking institution is not well established and lack academic rigor. Therefore, given that the transformation is spanning from the developed countries and well established in their

banks, this research examines the factors of digital transformation in Ghana – a developing country. These issues hinder the successful implementation of digital technologies. Moreover, given the complex nature of banking and electronic services, requiring managers and employees to adopt new behaviours due to technological infrastructure changes is not easy, especially in developing countries [4]. Therefore, it is important to examine the critical factors that hinder a banks intention to adopt digital transformation processes in a developing country context.

The paper first explores the current state of digital transformation in banks in Ghana during the COVID-19 era. It then provides strategies and potential solutions to a post COVID-19 era. Importantly, the author unearths the determinants of digital transformation adoption in banking institutions in a developing country context. In lieu of the COVID-19 crisis, the paper also provides a framework for emergency digital transformation to respond to future pandemics.

2. Literature Review

2.1. Digital Transformation Adoption

Digital transformation adoption (DTA) affects many dimensions of the bank, such as competitiveness, decision-making, strategic direction, productivity, business model, and customers [5]. The COVID-19 pandemic is driving financial institutions to rethink and adapt their banking strategies in order to stay operational. However, banking institutions in developed countries are likely to perform better in turbulent situations like the COVID-19 pandemic than their counterparts in developing countries because of the heavy investment in information technology that increases resilience [6]. For example, McKinsey & Company found out that developed Asian countries utilised technology quicker to COVID-19 pandemic disruption to assist contact tracing and inform customers of product and services offerings [7]. Thus, to understand digital transformation, it is important to examine the current state of banks in the COVID-19 era and determine factors that can shape the adoption of digital transformation in developing countries.

Table 1 shows selected studies on digital transformation adoption, factors and the approaches used to unearth the phenomenon in different contexts. As shown, most of the studies used a quantitative approach, thereby creating a gap in the qualitative approach in digital transformation. This study serves as an opportunity to contribute to both research methodologies to understand the phenomenon holistically from a developing country context.

Table 1. Selected Studies on Digital Transformation Adoption

Authors	Theory	Country	Methodology	Adoption Factors
[8]	UTAUT2	Jordan	Questionnaire completed by 500 bank customers Structural Equation Modeling	Hedonic motivation Self-efficacy Habit, Trust
[9]	UTAUTx Consumer-related constructs	United Kingdom	Questionnaire completed by 268 consumers Structural Equation Modelling	Performance expectancy Social influence Innovativeness Trust, Perceived risk
[10]	UTAUT	United Arab Emirates	Questionnaire completed by 638 citizens Exploratory Factor Analysis	Effort expectancy, Facilitating conditions Trust
[11	UTAUT	Switzerland	Questionnaire completed by 462 freshmen Partial Least Squares Structural Equation Modeling	Trustworthiness Social presence Adaptiveness Appearance
[12]	Technology Acceptance Model (TAM)	Greece	Questionnaire completed by 161 employees in the banking sector Multivariate Regression Analysis	Perceived ease of use Perceived usefulness Perceived self-efficacy
[13]	UTAUT2	India	Questionnaire completed by 568 rural women Structural Equation Modeling	Effort expectancy, Habit Facilitating conditions Perceived competence

2.2. The Importance of Digital Transformation

Transforming banks with technologies (also called digital transformation) is a holistic concept that looks at the relationship between technologies, organisations, and strategic changes [1]. The concept is driven by technology advancement – i.e. using digital, mobile, social and new technologies to change customers' expectations [14]. Nicknamed digital entrepreneurship, the term digital transformation is often understood as the upfront implementation of the latest technologies to enhance business. However, investing in technologies requires a clear understanding of the relationship between the organisational and technological culture and institutional resilience to anticipate, cope, adapt, and not only the risk aspect. This makes the term disruptive, transformative, and unpredictable [15,16]. In lieu of the transformation, banks are adopting customer-centric approaches rather than product-centric to remain competitive [17]. Identified areas for improvement include addressing neglected customer segments, reducing bank operating costs, firm assessment of the traditional and online banking, optimising business processes, understanding customer usage behaviour of bank resources, improving customer selection and investing in artificial intelligence and big data analytics [17,18].

Notwithstanding the novel idea of transforming banks with technology, some barriers do exist. Barriers identified in the literature include high acquisition cost, internet access, online banking platform, and customer ICT performance. The startup cost for electronic banking and implementation of digitalisation in banks may be expensive [18]. [19] found some barriers to include attitude towards change, the high cost in purchasing and cost of greenness at the organisational level, implementing and maintaining technology equipment affect the adoption of online banking. Hosting online banking for customers is affected by insufficient knowledge and limited funding resources [20]. Consequently, complementing face-to-face service with online banking may cost more to harness and develop.

Despite the several benefits of online banking to customers, research shows that the adoption of self-service technology is less patronised [21]. According to [22], customers' acceptance of technology, age structure and use behaviour affect their knowledge and expectation towards digital technology, performability and use. Other obstacles include the difficulty of use and lack of technology skills with electronic banking accounts for the constraints to digital transformation in banks.

2.3. Developing Country Context Scenario

In the Ghanaian banking sector, COIVD-19 was expected to create a fiscal gap of GHS 11.4 billion. The Government of Ghana put in the following measures a) improve the FX inflows and increase government debt by US$ 1 billion through the IMF rapid credit facility, b) defer interest payment on non-marketable instruments with an expected GHS 1.2 billion impact cashflows from investment instruments, and c) adjust expenditure on Capex and goods and services to a tune of GHS 1.2 billion [23]. The Bank of Ghana also put in measures such as policy rate and reserves, interest payment, commercial bank support, and support to mobile money users to curb the impact of the pandemic [24]. There were free mobile money transfers up to GHS100 (approximately 17 USD) and a 0.75% flat rate on mobile money transfers above GH₵100 (equivalent to 16 USD) championed by the Standard Chartered Bank of Ghana to support the digital transformation agenda. ABSA Bank Ghana also promoted waiver of charges on instant interbank transfers on its digital channels and free mobile money transfers up to GH₵100 to customers.

2.4. Conceptual Model and Hypothesis Development

As shown in Fig. 1 are the four determinants (price value, hedonic motivation, inherent innovativeness (IN) and technology readiness (TR) have been included in the conceptual model. This study introduces IN and TR in addition to [24] proposed and validated constructs (PV and HM). The motivation of the proposed conceptual model is from [24] unified theory of acceptance and use of technology (UTAUT2). IN and TR have been extensively used and noticed in the IS and marketing literature [25,45].

Hedonic motivation

HM is the feeling of excitement, cheerfulness and enjoyment towards the use of technology [24]. Prior studies have found that factors associated with HM such as enjoyment and excitement, are the essential predictors of users' intention to adopt a technology [4]. Further, [46] found the positive influence of HM on customers intention to use innovation and consequently enhance their intrinsic motivation. Therefore, the following hypothesis is deduced:

H1: HM influences the adoption of digital transformation.

Price value

Previous research has found PR's positive influence on institutions and employees' intention to use technology [24]. However, for a banking institution to consider the use of technology, certain consideration is made. For instance, when the advantage of acquiring and using a banking innovation is greater than the PV of the associated banking cost, the PV will positively influence intention to adopt. Therefore, the following hypothesis:

H2: PV influences the adoption of digital transformation.

Inherent Innovativeness

IN is the degree to which a bank has interest and curiosity in innovations related to banking processes. [41] indicated that consumers or institutions are considered innovative when they experiment with new things, ultimately leading them to adopt new technology solutions. Bank managers with innovativeness are more willing and inclined to try new technologies [26]. In this regard, a bank's inherent innovativeness, especially during the Covid-19 pandemic, will promote the adoption of new technologies. Therefore, the following hypothesis:

H3: IN influences the adoption of digital transformation.

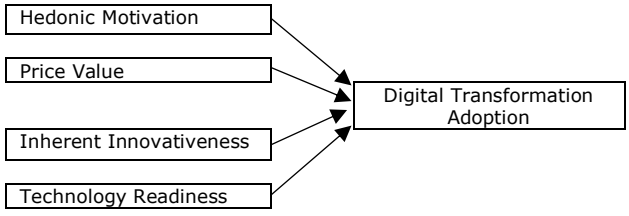


Fig. 1. Conceptual Model of Digital Transformation Adoption

Technology Readiness

Prior studies [25, 27, 28] have found that a firm's positive attitude towards new technologies may influence them to believe that the technology can improve the firm's efficiency, flexibility and business process continuity. Similarly, [27] found that TR influences customer satisfaction and intention to use new technologies. Therefore, the following hypothesis:

H4: TR influences the adoption of digital transformation.

3. Methodology

A qualitative and quantitative method was employed to determine digital transformation adoption in a developing country context through a structural equation modelling. Further, the study explores the impact of COIVD-19 on banking in Ghana and barriers to technological transformation in banking through an inquiry approach. In a qualitative study, the researcher learns from the participants through exploration. The critical phenomenon of this approach is the process, idea, and key concept studied [29]. A list of representations of the categories and case descriptions of COVID-19 impact on banks was formulated from extant literature based on participants' views. Data for the study was collected through online interviews (with a semi-structured guide) to adhere to COVID-19 protocols. Interviews allow the researcher to get an in-depth understanding of the phenomenon of study – feelings and respondents' attitudes [29]. The recurring themes from extant literature served as the research instrument for the analysis and were screened for validity by the researcher in Ghana.

Population and Sampling

All private banks in the Greater Accra Region of Ghana formed the target population. In all, twenty-eight banks were identified. Forty (40) of the managers of the banks were randomly selected to form part of this study. The managers came from the operations (16), general manager (4), client solution (4), electronic banking (10), SME (2), relationship (3) and corporate departments (1) in the bank (see Table 2). The majority of the positions were occupied by males (65%) while females occupied 35% (see Table 1).

Data Analysis

The qualitative analysis started by transcribing the recorded interviews using NVivo 11.0 software and analysing them into categories and sub-categories [30]. The software supports categorising unstructured data in the form of an audit trail to ensure coding and data interpretation consistency. The researcher replayed the interviews severally to ensure transcription accuracy, and interviews transcribed them verbatim. Pseudonyms were used for the participants to ensure confidentiality.

The quantitative analysis was assessed using the SmartPLS. The assessment of the model validation was performed in two steps, a) evaluation of the measurement model and b) examining the structural model by assessing collinearity issues and predictive relevance of the variables.

Table 2. Demographic data of respondents (n=40)

Demographic information	Frequency
Gender	
Male	26
Female	14
Position held	
Operations manager	16
General manager	4
Head of client solution	4
Electronic banking specialist	10
Relationship manager	3
Head, SME banking	2
Corporate manager	1
Years of banking service	
1-5 years	9
6-10 years	13
11-15 years	11
16 years and above	7
Did COVID-19 affect your bank operations	
Yes	40
No	0

Table 2 (continued)

Which delivery channel is the most preferred during the COVID-19 pandemic	
Internet banking and emails	9
Mobile banking	19
Automated teller	9
Face-to-face transaction	3

4. Results

The result of the study is in two parts. The first part focuses on the results of the qualitative analysis, while the part two focus on the quantitative analysis.

Part I: Qualitative Analysis

Barriers and Strategies of Digital Transformation

As shown in Table 3 are the identified themes on the impact of COVID-19 on banks and challenges associated with digital transformation as referred to in the literature and predominant in Ghana banks. The interviews suggest that all the banks in Ghana are affected by the COVID-19 pandemic and face many challenges migrating to digital transformation. The section concludes with post COVID-19 strategies identified in the study and supported by extant literature.

Table 3. Quantitative categories and sub-categories of the interview guide

Categories	Sub-categories
The impact of COVID-19 on banks	• Customer outcome • Social interaction • Traditional versus online banking
Barriers to digital banking transformation	• Cost • Customer • Internet access/exposure to risk
Post COVID-19 strategies to promote digital transformation in banking	• Fostering electronic banking acceptance • Employee • Deploying artificial intelligence • Harnessing the power of big data analytics

Source: Authors' categorisation

Current State of Banking in Ghana – The Impact of COVID-19 Pandemic

To understand the barriers that prevent digital transformation in the COVID-19 era, the study explored Ghana's current state of banking. Most of the managers interviewed indicated that the COVID-19 has negatively impacted their bank operations. However, the impact has opened an avenue for customers to patronise their online services (internet banking and mobile banking), which has been dormant since implementation. Table 4 is a detailed representation of the categories, sub-categories and case description of managers' experience with the impact of COVID-19 on banks in Ghana.

In conclusion, the managers revealed that the impact of the COVID-19 is the advent of new technologies and corporate change culture in banking. However, the change depends on employee and customer acceptance, understanding, and perception of digital transformation.

Table 4. Representation of the categories and case description of COVID-19 impact on banks

Main category	Sub-category	Case description	Number of sub-codes
Customers preparedness	Technology acceptance and innovation adoption during the crisis is low	Customers initially reject the change. However, technologically inclined customers get used to the shift knowing there is a limited option. Customers need to be involved with banks' transformation to encourage use.	8
	Usage behaviour	Prior knowledge of systems encourages usage behaviour. Customers gradual shift from analog to digital services is benefiting other online service use and acceptance. Also, the expectation and use of online services had a reflection on customers' ages.	6
	Expectations	The study revealed that customers' expectations are diverse. In contrast, some prefer to use a customer helpline, and others prefer instant messaging on our platform. Others chose to come to the banking hall, knowing the long waiting time to address issues that can be done on the phone.	4
	Traditional versus online approach	The interviews revealed that online services were fast and the best while most preferred personal face-to-face banking as the most suitable and secured. Some customers refused to obey the social distancing protocol despite signage	3
	Internet access and network instability	Customers complained of poor network slowing their business activities with the bank. Interesting, much focus was on the premier customers.	9
	Low deposits	Drop-in daily transactional limits, coupled with low income and deposits (heightened credit risks). Most clients were not able to honour their monthly loan payments	3
Employees/ Bank preparedness	High Cost	Going online is very expensive. Most employees were provided with laptops and internet access to work from home.	10
	Technology acceptance and innovation adoption during the crisis is low	Acceptance of working remotely with technologies was fraught with low interest by employees. However, initial training on software and hardware encouraged acceptance and involvement.	6
	Technological constraints and internet instability	Aside from the high cost of digitalisation and IT equipment provided to employees to work remotely, there were constant internet instabilities. Most were staying in the hinterlands and remote areas. Some employees were also overwhelmed by the technological work from home policy. Working remotely with different applications and devices was a challenge for some.	7
	Distractions	Working from home requires employees to have high self-discipline. Distraction from family members, home electronics, and social media were some of the employees' challenges.	5

Source: Authors' representation

"We have no option now but move all processes and operations online and train employees to develop the skillset to work in a digital world." [Kwesi – Manager]

"Global Covid-19 had a very negative impact, but the banking sector has greatly benefited from COVID-19 in the sense that it has pushed banks to be innovative and serve their customers beyond the traditional way of running a business. Banks have become more receptive in terms of technology." [Serwaa – Manageress]

"Most training is done online, and it is a good thing. Also, there was easy adoption of internet banking and marketing communication during the pandemic. Thus, more customers have switched to the use of online banking, which will help the bank save on cost." [Mich – Manager]

"Covid has exposed how vulnerable most financial institutions are. Most banks were caught unprepared during the lockdown period, and technological innovations will ensure

convenient banking customers. In any way, there was easy adoption of internet banking and marketing communication during the pandemic." [Ellen – Manageress]

The interview revealed that managers need to develop a Quick Pandemic Recovery Plan and intensify digital channels' use. Promote remote onboarding of customers and focus on transactional businesses. The following were managers take-home experiences with the COVID-19 pandemic in Ghana:

1. Digital drive
2. Process optimisation and customer onboarding
3. Equipping all colleagues to work remotely
4. Video conferencing and focus on technological constraints

Barriers to Digital Transformation

Cost. Among the significant challenges mentioned by the managers, the cost was the most recurring theme that prevented most banks in Ghana from taking an active part in electronic banking. According to the managers, complimenting face-to-face business with electronic banking is very expensive.

"Digital technologies and infrastructure to meet banking transformation are associated with high costs and need to be managed well so that much expenses are not shifted to the customers. The cost is a major hindrance to digital banking transformation." [Akosua – Manageress]

"To go online is very expensive, especially in this COVID-19 period. When the president announced the lockdown, my bank decided to secure remote working systems. The systems were very expensive if we have continued to work from the office. Still, the benefit was good, notwithstanding the new modalities for restructuring credit facilities for customers and grant moratorium." [Joseph – Manager]

"There is no readily available public funding for implementation of banking technologies. Most people assume that the banks are self-reliant and have the necessary resources and financial powers to acquire, implement and maintain technology and banking transformation without support. Even so, customers barely patronise the technology services when implemented. Therefore, the cost becomes double." [Eunice – Manageress]

Customers. The managers revealed that customers' readiness to accept new technologies and the overall digital transformation is challenging. However, managers reveal more transactions from customers' bank accounts to their mobile money wallet, indicating the trend of future payment (i.e., mobile) methods during a crisis. The banks also showed low adoption of innovation and difficulty adjusting to new technologies during the crisis, especially remotely working from home. In short, there was low technology acceptance and innovation adoption during the COVID-19 pandemic among customers.

"Except for customers in the big cities, customer technology acceptance, use and trust in our digital services are low. We need to do more sensitisation and education for customers to prevent the rejection of digital technologies in banking because it will be the medium of service and transaction for the day. Nevertheless, it is important to say that customers' behaviour gradually changes from analog thinking to digital use. With the launch of Bank of Ghana's mobile money interoperability system, most customers expect to use digital skills to make financial transactions without hassle." [Enoch – Manager]

"There is a paradigm shift with our digital banking transformation. We have realised that the young are more ready and skilled, and most opt for digital services. Age has therefore become a determinant with customers knowledge and expectations towards digital technologies in banking." [Sandra – Manageress]

"Some customers expect the digital transformation of their banking activities. However, the banks implement these technologies, customers barely use them conveniently or to their full advantage. Some customers even prefer coming to the banking square to confirm their transaction after conducting it online." [Rich – Manager]

"I would say the lockdown was the beginning of contactless banking. Since the president announced it, digitalisation has been the order of the day. This led to some bank closure and

face-to-face interaction and transactions. Consequently, the action affected our financial inflows – low transactions." [Kofi – Manager]

"Oh, on the employee side, the acceptance of innovation was fraught with challenges especially training them on new technologies to enable their work from home policy. Some were already tech-savvy, but others took a while to come to the new normal reality. They struggle to adjust to new technologies, especially those in our satellite and remote/rural parts of Accra." [Claudia – Manageress]

Exposure to risks. The interviews revealed that banks' digital transformation exposes the customer and the bank to online risks. The [31] disclosed 2295 fraud cases in 2019 compared to 2,175 cases in 2018. As reported by a manager;

"In our effort to adopt latest banking technologies have made our banks more vulnerable to attacks and risk such financial scamming, anti-money laundering, identity theft and more sophisticated types of cyber-crime. This attitude discourages customers from adopting online services entirely." [Eugene – Manager]

Prevalent fraud cases identified in the Ghanaian banking industry include suppressing cash and deposits, forgery and manipulation of documents, cheque fraud and cyber and email fraud. As banks continue to go fully digital, they are strongly advised to enhance monitoring processes, both physical and technological and cybersecurity infrastructure to reduce the fraudulent incidence of call diversion and prevent unauthorised access in a post COVID-19 era.

"Automation of banking process could lead to job loss and increase fraud activities due to high use of online banking. There is a lot to think about with full automation." [Yaw – Manager]

Internet access and digital means Managers persistently lamented the poor internet accessibility by customers in accessing digital services. The interviews revealed that the number of customers that access the online services was comparably low to the total customer base.

"Oh ok. You know that internet access is a huge problem in our part of the world not to talk of the customers' digital means (to acquire smartphones and laptops) to access our online services. Presumably, most banks know the low number of customers that will use their digital services; however, we have no option but to implement. I know the new normal of doing business will force customers to online services." [Freda – Manageress]

"One major challenge of customers accessing our online services is the means. Smart technologies and engaging in online activities are quite expensive." [Josephine – Manageress]

Strategies to Promote Digital Transformation in Post COVID-19

The managers were asked to state post COVID-19 strategies that can promote digital transformation in banks. After the researcher examined the strategy, the theme was categorised into two as supported in extant literature reviewed. The first category includes a) fostering online banking acceptance, b) digital literacy for customers, and c) hybrid banking. These factors contextualise the Ghanaian bank's experience and offer post Covid-19 strategies to promote digital transformation in developing countries. The second includes a) deploying artificial intelligence and machine learning, b) harnessing the power of big data analytics, and c) addressing the digital transformation myth. These factors focus on providing recommendations on how new technologies can transform banks to adapt to the new normal.

Fostering online banking acceptance. The interviews revealed that customers are often slow to embrace new banking technologies because of security reasons. Banks today cooperate with customers to know their challenges and find

solutions to promote digital corporate transformation. One manager was of the view that *"…. putting measures such as improving customers self-efficacy with ease-of-use technologies would increase online banking acceptance. Engaging customers in a recreational and marketable fashion to try digital technologies would also encourage acceptance" [Kate – Manageress].* According to [32], collaborative online activities and prompt feedback can promote exchange experiences and ideas among customers and banks, promoting competence and increasing users' self-efficacy. Tailoring systems to meet user characteristics and customers' competence may ensure rapid acceptance of online banking. This leads to a user-need analysis for adaptive electronic banking to meet the customers' expectations (personalisation of systems) [33].

Further, the interviews revealed that technological conditions necessary for digital transformation had been delayed and prerequisites not met. "*With my several years in banking, one common problem is the will and implementation power to see through digital projects. When my bank decided to connect all its satellite banks, I remember we secured the infrastructure and started creating for each satellite. However, out of the 34 satellite banks, only six (6) were connected. On the bank side, we were not ready – financially, human capital, technologically and infrastructure. Waste of resources, I guess, because we will need enhanced servers and infrastructure because of time-bound factor." [Dan – Electronic Banking Specialist]*

Hybrid banking. The interviews revealed that Ghanaian customers are yet to come to terms with online banking to the full extent. The implication is that a "hybrid banking" approach that is face-to-face interaction and online service banking will play an essential role in the long-term survival of banks in developing countries. Till customers embrace the entire digital transformation, banks should reduce the heavily and unprepared cost associated with technologies and approach it from the gradual user acceptance stage. The COVID-19 has proven that customers are dynamic but approach systems and regulations in a gradual process. Systems acceptance has been based on customer preparedness, self-efficacy, secure banking, and prior knowledge. Hybrid banking will also promote equity in banking, which are essential responsibilities of financial and banking institutions.

Deploying artificial intelligence. The interviews further revealed artificial intelligence and its important role in the future development of banks in Ghana. One interviewee stated how they have started collaborating with universities and research institutions for strategic support and new infrastructure development and application to advance AI. *"One essential resource driver available to customers is their level of comfort and ease of use with new information technologies. Artificial intelligence is permeating banking operations and such we have collaborated with a university to provide better insight in its application and integration with our current systems" [Kwame – Manager]*. [34] suggest that "technology integration is a critical component of value co-creation because of its ability to integrate, collaborate and access other resources." Reshaping service delivery and technology's ability to co-evolve business processes is more evident with artificial intelligence than any known technological forces. Research has found that AI is intertwined with customers' perception and technology acceptance, readiness, trust, and security [35]. The efficient use of AI in a higher-value use context, such as real-time personalised investment portfolios and real-time personalised advice on bank accounts, will provide seamless opportunities and insight for banks in emerging markets [36]. Recent utilisation of AI-enabled devices in banking, coupled with customers growing interest and adoption of AI technologies, allows collection, analysis, storage and access of large datasets, both by financial institutions and technology providers [37].

Harnessing the power of big data analytics. The new normal requires banks to spot and anticipate recent trends, demands, and requirements within the shortest possible time, and this is possible with data analytics. This move towards new customer insight and services approaches may lead to a paradigm shift where banks invest. One interviewee confirms the emergence and importance of big data analytics and how the organisation is actively looking into such investment. "*I think the next big thing happening to banks is the advent of big data. Big data analytics has already been used to evaluate financial crime management solution rules, customer-centric analytics and micro-segmentation, IVR analysis, and B2B merchant insights. One cannot afford to miss out on its potentials for bank revolution." [Prince – Manager].* By harnessing the power of data analytics, a business can provide opportunities and options for growth. The emergence of COVID-19 has shown that data analytics play a predominant role in business anticipation, coping and recovery, and can enhance the agility of organisations [38].

Addressing digital transformation myth. The interviews revealed how digital transformation seems to be on the minds and agenda of electronic banking specialists and operation managers. Indicating they are already in touch with tech giants to transform business operations because of the COVID-19 lessons. *"The little COVID-19 has shown us is enough to go fully digital and transform our processes" [Jona – Operations Manager].* However, the digital transformation migration is a whole cooperative adaptation process. As stated by [39], "the application of these new technologies and their appropriate implementation to improve business performance is an important issue to the organisation" and if not well implemented, would affect all dimensions of the organisation, including business model, competitiveness, decision-making, productivity, strategic direction, and customers [40].

Consequently, preparing a banking institution after the COVID-19 pandemic involves more than acquiring the services of a tech giant or moving information to the cloud. It involves embracing technology to connect the business end-to-end, i.e., the employees, suppliers, stakeholders, and customers, and moving the business process to the next level. By leveraging technology in the various aspect of banking, the electronic specialist or IT lead can approach digital transformation as such; a) tackle digital transformation in a phased approach, e.g., the financial aspect versus the e-commerce aspect, b) keep employees and customers at the centre of the transformation, and c) digital transformation should be the culture and way of life of the bank.

Part II: Quantitative Analysis

Determinants of Digital Transformation Adoption

This study adopts a two-stage approach to examine the determinants of digital transformation adoption, i.e. measurement model assessment and structural model assessment. The measurement items of price value and hedonic motivation were adapted from [24], inherent innovativeness was adapted from [41], and technology readiness was adapted from [25]. All the measurement items were on five Likert scales.

Measurement Model Assessment

The assessment of the construct reliability was conducted via composite reliability (CR) and average variance extracted (AVE). As shown in Table 4, the CR values range from 0.839 to 0.909, indicating that all the latent constructs met the minimum threshold of 0.70. However, the assessment of the Cronbach alpha indicates that all the latent construct values exceeded 0.7 thresholds apart from PV (0.683)

[42]. Similarly, the AVE all exceeded the minimum threshold of 0.5 [44]. Further, the estimation of collinearity was assessed using variance inflation factor (VIF). As shown in Table 5, all the values were less than 5, indicating no collinearity issue [43].

Table 5. Construct Reliability and Validity

Construct	CR	rho_A	AVE	CrA	VIF	R2
Price value	0.853	0.857	0.746	0.683	2.341	
Hedonic motivation	0.839	0.791	0.638	0.721	1.092	
Inherent innovativeness	0.845	0.778	0.649	0.728	2.664	
Technological readiness	0.842	0.728	0.640	0.722	1.807	
Digital Transformation Adoption	0.909	0.849	0.768	0.849		0.803

The model's predictive capability was also assessed using the R2 through a bootstrapping procedure. R2 values 0f 0.25, 0.50 and 0.75 can be classified as weak, moderate, and strong, respectively [43]. From the analysis, the R2 value is 0.803, indicating a strong predictive power of the exogenous constructs. Thus, explaining 80.3% of the variance in digital transformation adoption. The discriminant validity was examined to establish the constructs' validity using the Fornell-Larcker, as shown in Table 6 [44].

Table 6. Discriminant Validity – Fornell-Larcker

	DTA	HM	IN	PV	TR
DTA	0.876				
HM	0.239	0.799			
IN	0.839	0.186	0.805		
PV	0.729	0.281	0.736	0.864	
TR	0.768	0.196	0.654	0.568	0.800

Structural Model Assessment

Findings from the structural model estimation indicate that three hypotheses (H1, H3 and H4) in the proposed conceptual model were supported, as shown in Table 7. The effect size (F2) also indicates that 0.02, 0.15 and 0.35 represent small, medium and large values, respectively.

Table 7. Path Coefficients

Path	Coefficients β	P Values	STDEV	F2	Support
Price Value → DTA	0.166	0.027	2.225	0.059	*H1:* Supported
Hedonic Motivation → DTA	0.033	0.488	0.694	0.005	*H2:* Not supported
Inherent Innovativeness → DTA	0.479	0.000	5.653	0.436	*H3:* Supported
Technology Readiness → DTA	0.354	0.000	4.383	0.352	*H4:* Supported

Having identified the barriers and strategies of digital transformation in a developing country context like Ghana, the present study represents an important contribution to the factors that determine the adoption of digital transformation in banks. Basically, price value significantly predicted digital transformation adoption in banks. Moreover, inherent innovativeness and technology readiness significantly predicted the adoption of digital transformation in a developing country. Conversely, hedonic motivation did not influence banks decision to adopt.

As shown in Fig. 2, inherent innovativeness seems to be the strongest predictor of digital transformation adoption. This means that banks in developing countries should consider customers and employees' curiosity and interest in trying new banking technologies. Importantly, this prediction is followed by technology readiness. Thus, a positive banks attitude towards new banking technologies will influence customers and employees in the digital transformation agenda. This finding is similar to [27] study on customer satisfaction and interest in using new

technologies. Consequently, banks should focus on developing employees and customers' competence (self-efficacy) in making this a reality in developing countries.

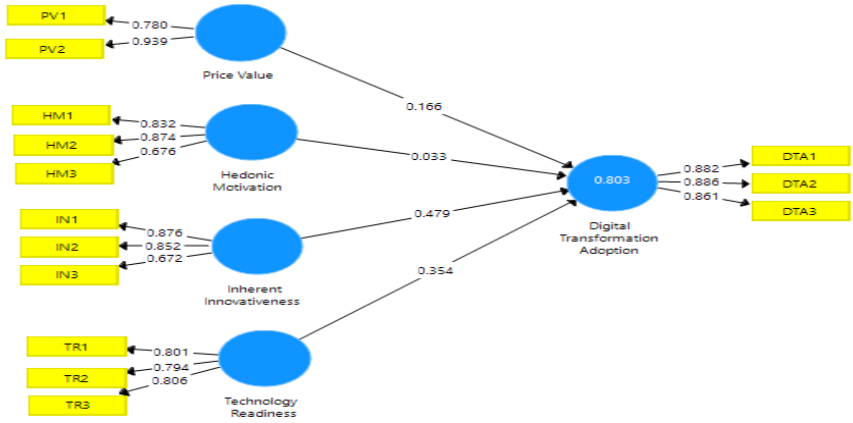


Fig. 2. Results of the research model (constructs adapted from [24, 25, 41])

The non-significant influence of hedonic motivation on digital transformation adoption is not consistent with results from previous studies [8]. However, many reasons can be assigned. The first is the rate at which the customers and banks' price value and technological readiness surpass the excitement and fun in acquiring the digital technologies. As shown in Figure 2, price value was the third strongest factor predicting digital transformation adoption. This means that banks do prioritise the cost as paramount to hedonic motivation. Nevertheless, the non-significance of hedonic motivation among the four constructs does not make it less valuable.

A conceptual model for emergency digital transformation to respond to future pandemics

As shown in Fig. 3, the model serves as a framework for leveraging digital technologies in banking and financial institutions to respond to future pandemics such as the novel COVID-19. The framework is based on the validated proposed model in Figure 2. The model suggests that for successful implementation of digital transformation in banking, financial institutions must first promote and incentivise digital technologies to minimise person-to-person contact and reduce the spread of the virus (inherent innovativeness). Also, there should be a secure, resilient digital technologies for the robustness of platforms and systems to facilitate financial transactions, i.e., mobile money transfer in developing countries during pandemics (price value). For the purpose of business continuity, firms should enable regulations to suit prevailing circumstances and context of banking which may boost agent and merchant operations (technology readiness) [25], especially in developing countries like Ghana where mobile money transfer surpasses all payment platforms.

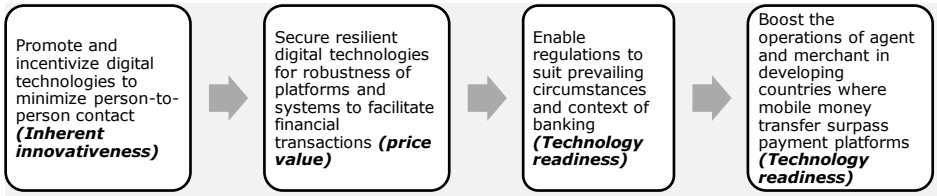


Fig. 3 Proposed conceptual model for emergency digital transformation

Contribution to research, policy and practice

This study makes important contributions to research, policy and practice to organisations that want to carry out digital initiatives. In terms of research, this study contributes to the body of knowledge on digital transformation adoption and bank transition to digitalisation by revealing the possible challenges and solutions that potential banks will face in a developing country context. It also tested and validated PV, IN and TR as key determinants of digital transformation adoption in a developing country. Given the existence of societal and cultural peculiarities in different nations, this contribution is important. This study also provides a framework that future banking and financial institutions can adopt to tackle pandemics and reduce the impact on the economy.

Concerning practice, managers of banking institutions can draw on the experiences and factors likely to affect their transition to digitalisation from this study. Thus, banking institutions venturing into digital transformations better understand challenges and strategies, a knowledge gap arguably not readily available in post Covid-19 to banking institutions in developing countries. Therefore, banking and financial institutions planning on digital transformation in developing countries need to consider fostering online banking acceptance, hybrid banking, deploying artificial intelligence, harnessing the power of big data analytics and addressing the digital transformation myth. The study contributes to policy by drawing practitioners' attention to the fact that creating a favourable banking environment enabled by digital technologies will positively influence the adoption of online banking services.

5. Conclusion and future research

Despite banks' huge investment in digital technologies in developing countries like Ghana, challenges such as technology acceptance, online service reluctance, and poor usage behaviour persist in the banking sector, especially when COVID-19 protocol encourages use. In lieu of the COVID-19 pandemic on the global market, the study examined the impact on banks in Ghana and revisited how digital technology is fundamental to online banking and services. This further led to exploring the barriers to the digital transformation taking place in developing countries like Ghana. The study findings suggest that digital technologies use is not entirely accepted by customers and employees, and hybrid banking is effective as banks continue their digital transformation. On the customer side, technological constraints and internet instability during the COVID-19 period prevented successful financial transactions. Employees' reluctance to work remotely, difficulty adjusting to new technologies, distraction from family members and home electronics, technological constraints, and internet instability were the perceived challenges associated with the COVID-19 pandemic.

Further, this study emphasises the point that digital transformation adoption cannot be studied from one source or context alone. Rather, the contextual price value or cost, the level of inherent innovativeness and the technological readiness of banks should be considered as they may inhibit or promote digital transformation adoption in developing countries.

Managers' perspectives on barriers to digital transformation in banking include cost, customers' readiness to accept and use digital technologies, exposure to risks, and customers' internet access and digital means. Post COVID-19 strategies to promote digital transformation in banking in developing countries entail fostering online banking acceptance, adopting a hybrid banking approach, investing in artificial intelligence, and harnessing big data analytics. Also, managers can tackle digital transformation in a phased approach by keeping employees and customers at the center of the transformation and making digital transformation the bank's culture and way of life.

The participants' interest in this study underscores the importance of the exploratory approach to digital technologies in banks, notwithstanding the need to explore other methodologies to garner new insights. Importantly, participants' direction to digital transformation in banks was not straightforward, revealing that the field of research is still young, and much attention is needed in the developing country context. Future research can begin by exploring the perceived and actual responses to the use and implementation of digital transformation. Lastly, the starting place for future research is to empirically explore the conceptual model for emergency digital transformation in responding to future pandemics. New insight from extant literature can broaden the model limitation's scope for better preparedness to pandemics in the banking sector.

The study has some limitations. First, it is difficult to generalise the study because the sample is very small. The study only focused on the managers' perspectives of the impact of COVID-19 on banks in Ghana and the technological transformation of banks without including the voices of customers and employees during the pandemic. Further, the study participants are from private banks in Ghana without including public banks and other financial institutions like microfinance impacted by the COVID-19 and experiencing a digital transformation**.**

References

1. Matt, C., Hess, T., Benlian, A.: Digital transformation strategies. *Business & Information Systems Engineering*, *57*(5), 339-343(2015).
2. Wessel, L., Baiyere, A., Ologeanu-Taddei, R., Cha, J., Blegind-Jensen, T.: Unpacking the difference between digital transformation and IT-enabled organisational transformation. *Journal of the Association for Information Systems*, *22*(1), 102-129 (2021).
3. Avgerou, C.: The significance of context in information systems and organisational change. *Information systems journal*, *11*(1), 43-63(2001).
4. Curran, J. M., Meuter, M. L.: Encouraging existing customers to switch to self-service technologies: put a little fun in their lives. *Journal of Marketing Theory and Practice*, *15*(4), 283-298(2007).
5. Cohen, B. Amorós, J.E., Lundy, L.: The generative potential of emerging technology to support startups and new ecosystems. *Business Horizon. 60*(2017).
6. Chung, T. T. R., Liang, T. P., Peng, C. H., Chen, D. N., Sharma, P.: Knowledge creation and organisational performance: moderating and mediating processes from an organisational agility perspective. *AIS Transactions on Human-Computer Interaction*, *11*(2), 79-106 (2019).
7. Choudhury, S. R.: How technology helped Asian countries cope in the coronavirus pandemic. CNBC Online (2020). https://www.cnbc.com/2020/05/13/coronavirusmckinsey-says-technology-helping-asia-cope-better.html, last accessed 2021/05/13.
8. Alalwan, A. A., Dwivedi, Y. K., Rana, N. P., Lal, B., & Williams, M. D.: Consumer adoption of Internet banking in Jordan: Examining the role of hedonic motivation, habit, self-efficacy and trust. *Journal of Financial Services Marketing*, *20*(2), 145-157 (2015).
9. Slade, E.L., Dwivedi, Y.K., Piercy, N.C., & Williams, M.D.: Modeling consumers' adoption intentions of remote mobile payments in the United Kingdom: extending UTAUT with innovativeness, risk, and trust. *Psychology & Marketing*, *32*(8), 860-873(2015).
10. Al Mansoori, K. A., Sarabdeen, J., & Tchantchane, A. L.: Investigating Emirati citizens' adoption of e-government services in Abu Dhabi using modified UTAUT model. *Information Technology & People*. 31 (2), 455-481 (2018).
11. Guggemos, J., Seufert, S., Sonderegger, S.: Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology*, *51*(5), 1864-1883(2020).
12. Kitsios, F., Giatsidis, I., Kamariotou, M.: Digital Transformation and Strategy in the Banking Sector: Evaluating the Acceptance Rate of E-Services. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(3), 204(2021).
13. Manrai, R., Goel, U., & Yadav, P.D.: Factors affecting adoption of digital payments by semi-rural Indian women: extension of UTAUT-2 with self-determination theory and perceived credibility. *Aslib Journal of Information Management*. 73 (6), 814-838(2021).

14. Terrar, D.: What is Digital Transformation? Available online: http://www.theagileelephant.com/what-is-digitaltransformation/ (accessed on 2021/08/10).
15. Duchek, S.: Organisational resilience: a capability-based conceptualisation. *Business Research*, *13*(1), 215-246 (2020).
16. Krasonikolakis, I., Tsarbopoulos, M., Eng, T. Y.: Are incumbent banks bygones in the face of digital transformation? *Journal of General Management*, *46*(1), 60-69(2020).
17. Lotriet, R., Kokotwane Dltshego, K.: An assessment of perceptions concerning digital transformation at a South African commercial bank-a case of Anthropocene denial for the economy?0 *Tydskrif vir Geesteswetenskappe*, *60*(3), 687-707(2020).
18. Mărăcine, V., Voican, O., Scarlat, E.: The Digital Transformation and Disruption in Business Models of the Banks under the Impact of FinTech and BigTech. In Proceedings of the *International Conference on Business Excellence*; Walter de Gruyter GmbH: Berlin, Germany, 2020; pp. 294–305.
19. KamalulAriffin, N.S., Khalid, S. N.A., Wahid, N.A.: The barriers to the adoption of environmental management practices in the hotel industry: a study of Malaysian hotels. *Business Strategy Series*.*14*, 106-117(2013).
20. Vikneswaran, N., Anantharajah, S.A.: Green makeover for our hotels? *Q. DOE Update Environment Development and Sustainability. 2*, 10–12 (2012).
21. Estrella-Ramon, A., Sánchez-Pérez, M., Swinnen, G.: How customers' offline experience affects the adoption of online banking. *Internet Research. 26 (5),1072-1092* (2016).
22. Diener F, Špacek M.: Digital Transformation in Banking: A Managerial Perspective on Barriers to Change. *Sustainability* 13, 2032 (2021).
23. KPMG Ghana Report: Economic Impact of COVID – 19 on Ghana's banking industry (2020).https://assets.kpmg/content/dam/kpmg/gh/pdf/gh-KPMG-Advisory-Economic-Impact-of-COVID-19-on-Ghana%E2%80%99s-Banking-Industry.pdf, last accessed 2021/06/09.
24. Venkatesh, V., Thong, J. Y. L. and Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157-178.
25. Parasuraman, A.: Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies. Journal of Service Research, *2*(4), 307-320(2000).
26. Massey, A.P., Khatri, V., Montoya-Weiss, M.M.: Usability of online services: The role of technology readiness and context. *Decision Sciences*, *38*(2), 277-308(2007).
27. Lin, J.S.C., Hsieh, P.L. The influence of technology readiness on satisfaction and behavioral intentions toward self-service technologies. C*omputers in Human Behavior*, *23*(3), 1597-1615(2007).
28. Erdoğmuş, N., Esen, M.: An investigation of the effects of technology readiness on technology acceptance in e-HRM. *Procedia-Social and Behavioral Sciences*, *24*, 487-495 (2011).
29. Creswell, J. W.: Research design: qualitative, quantitative, and mixed methods approaches (5th ed.). Los Angeles: SAGE. (2018).
30. Miles, M.B., Huberman, A.M.: *Qualitative Data Analysis: An Expanded Sourcebook*, SAGE Publications, Thousand Oaks, CA. (1994).
31. Bank of Ghana.: The 2019 banking industry fraud report. https://www.bog.gov.gh/wp-content/uploads/2020/08/SUMMARY-OF-2019-FRAUD-REPORT.pdf, last accessed 2021/07/3.
32. Ngai, C.S., Lee, W.M., Ng, P.P., Wu, D.D.: Innovating an integrated approach to collaborative eLearning practices in higher education: The case study of a corporate communication e-platform. *Studies in Higher Education*, 44(11), 1990–2010 (2018).
33. Morrison, D.: E-learning strategies: How to get implementation and delivery right first time. England: John Wiley & Sons Ltd. (2003).
34. Zainuddin, N., Tam, L., McCosker, A.: Serving yourself: value self-creation in health care service. *Journal of Services Marketing*, 30(6), 586-600 (2016).
35. Shankar, A., Kumari, P. (2016). Factors affecting mobile banking adoption behavior in India. *The Journal of Internet Banking and Commerce*, *21*(1).
36. Payne, E.H.M., Dahl, A.J., Peltier, J.: Digital servitisation value co-creation framework for AI services: a research agenda for digital transformation in financial service ecosystems. *Journal of Research in Interactive Marketing*. 15 (2), 200-222(2021).
37. Moro, S., Cortez, P., Rita, P.: Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, *42*(3), 1314-1324(2015).
38. Gerald, E., Obianuju, A., Chukwunonso, N.: Strategic agility and performance of small and medium enterprises in the phase of Covid-19 pandemic. *International Journal of Financial, Accounting, and Management*, *2*(1), 41-50(2020).

39. Ismail, M. H., Khater, M., Zaki, M.: Digital business transformation and strategy: What do we know so far. *Cambridge Service Alliance*, *10*(2017).
40. Chanias, S., Myers, M. D., & Hess, T.: Digital transformation strategy making in pre-digital organisations: The case of a financial services provider.' *Journal of Strategic Information Systems*, 28, 17–33 (2019).
41. Shih CF, Venkatesh A.: Beyond adoption: Development and application of a use-diffusion model. Journal of marketing. 68(1):59-72 (2004)
42. Gefen, D., Straub, D., Boudreau, M. C.: Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems*, *4*(1), 7(2000).
43. Hair, J.F., Black, W.C., Babin, B.J. Anderson, R.E. *Multivariate Data Analysis*, 7th ed., Prentice Hall, Upper Saddle River, NJ. (2010).
44. Henseler, J., Ringle, C. M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135 (2015).
45. Anwar, A., Thongpapanl, N., & Ashraf, A. R.: Strategic imperatives of mobile commerce in developing countries: the influence of consumer innovativeness, ubiquity, perceived value, risk, and cost on usage. *Journal of Strategic Marketing*, *29*(8), 722-742(2021).
46. Riffai MM, Grant K, Edgar D. Big TAM in Oman: Exploring the promise of on-line banking, its adoption by customers and the challenges of banking in Oman. *International Journal of Information Management*. 32(3):239-50 (2012).

The Use of Digitization in Small and Medium-Sized Agricultural Enterprises: Evidence from the Czech Republic

Martina Valentová[1][0000-0002-4989-0281],
Lilia Dvořáková[1][0000-0001-6389-381X]

[1] University of West Bohemia, Faculty of Economics, Department of Finance and Accounting, Univerzitní 8, 301 00 Plzeň, Czech Republic

https://doi.org/10.33847/2686-8296.3.2_6

Received 10.10.2021/Revised 08.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. This paper presents the results of the research focused on the issues of digitization of agriculture in the context of current global developments and on the need to increase the amount of food produced in order to feed the continuously growing world population. The paper aims to analyze the use of the selected digitization tools in the crop production in the Czech Republic by small and medium-sized enterprises and to evaluate that from the economic point of view. To achieve the objectives, the desk research, the analysis and evaluation of secondary sources, and the method of directed interviews with managers and employees of the selected business entity were used. The research results evidence the positives and negatives of the use of the selected digitization tools within the crop primary production and represent the basis for further research aimed at the development of economic and financial management of small and medium-sized agricultural enterprises in the context of Agriculture 4.0 concept.

Keywords: Digitization, Drones, GPS, Smart agriculture, Agriculture 4.0.

1. Introduction

Agriculture has been accompanying mankind virtually since time immemorial and the last decades of social development have been associated with a massive technological development. Digitization has gradually permeated all fields and has not avoided the traditional sector of economy, that is agriculture.

In the 1980s, the concept of "precision agriculture" emerged, that is, farming with a focus on the individual approach to every single field and crop, with an effort to reduce the environmental impact of the activity itself [1, 2]. However, the then level of technology development hindered any major application in practice. With the further development of technologies (especially satellite navigation, the Internet and computer technology), many of the ideas of "precision agriculture" could have been implemented in the practical life of farmers. It is this introduction of 'smart technologies' that is often referred to as 'smart' or, as in the case of industry, '4.0'. Agriculture 4.0 aims to increase the use of robots and autonomous systems, and the extent and quality of data collection, processing, and their subsequent use (specifically GPS, drones, sensors, detectors, specialized software, etc.). Three principles underpin Agriculture 4.0: 'people, production and planet' [1, 3, 4, 5, 6, 7].

This article firstly shortly precents actual knowledge and studies focused on theoretical and practical implementation of digitalization principles in agriculture and briefly introduce situation in the Czech Republic focusing on small and medium-sized enterprises (then abbreviated SME).

Secondly presents methods which were used in data collection for this research. Results part is focused on using drones and GPS technology and their positives and negatives in practically agriculture (the crop production). A final summary follows with

limitations of research and opportunities of further research practical application of principles of Agriculture 4.0 in small and medium entrepreneurs.

The main objectives of the research are as follows:

- To evaluate the positives and negatives of using drones and GPS in crop production in the agricultural SMEs.
- To analyze the economic intensity of using drones and GPS technology in SMEs.
- To propose a reference costing for the acquisition of a drone in crop production.
- To discuss the barriers to the development of Agriculture 4.0 in SMEs.

The paper seeks to answer the following research question: "What technologies can Czech farmers use to obtain data on the crop development in crop production?" and also "Which of these potentially usable technologies is the most suitable and affordable one for a small agricultural entity?"

2. Literature Review

The issues of Agriculture 4.0 and digitization in the agricultural sector are addressed in general by, for example, Bollini et al. [5] or Zhai et al. [8]. This topic is being pursued across the entire world, as evidenced by the study of Kilavuz & Erdem [9], which examines the topic in Turkey, or, for example, the study of Luu et al. [10] from Vietnam. Agriculture 4.0 aims, among other things, to increase the production of agricultural commodities by 50% by 2050 in order to feed the planet´s ever-growing population. One concrete step is to extend digitization in agriculture, which means, for example, making use of satellite navigation systems to measure crop positions, the use of navigation in machines, the use of fertilizer application machines, automating harvesting operations and more. The overall aim is to reduce the workload of human machine operators and the associated error rate and diseconomy [11].

How do digitization trends specifically manifest themselves in the Czech agriculture in small and medium-sized enterprises? The Association of Small and Medium-Sized Enterprises and Sole Traders of the Czech Republic (abbreviated as AMSP CR in Czech) conducted a survey which shows that 8 out of 10 respondents can imagine specific elements and activities under the concept of agriculture digitalization. The research also shows that only 20% of the surveyed small and medium-sized Czech agricultural entities invest in robotic automation and digitization, and the same is true for production automation.

Another finding is that 64% of the agricultural enterprises surveyed use modern navigation systems, 59% use mapping systems, 54% use weather stations, and only 9% of farmers use drones. Based on the same survey, corporate entrepreneurs - corporate bodies - are more likely to invest in modern technologies than small entrepreneurs - natural persons. The survey was carried out among Czech small and medium-sized agricultural enterprises and self-employed persons [12]. The digitization of agriculture in the Czech Republic consists mainly in the use of GPS technologies or "smart machines" (e.g., digital milking, mapping of fields using satellites or aerial photographs, weather stations, sensors, etc.).

The benefits of these technologies in the agricultural sector include a reduction in the number of workers needed (In general, agriculture faces low employment attractiveness, especially among the younger generation.), a reduction in physical work, an increase in the volume of production and its quality, or an increase in the quality of data that are used as a basis for company decision-making [11, 13].

3. Data and Methodology

In order to achieve the set objectives, mainly the methods of qualitative research were used. In the first phase of the research, the desk research method was used to analyze and evaluate the secondary specialized, legislative, and statistical sources. In

addition, the methods of interpretation and comparison were used. The empirical part of the research was carried out using the method of directed interviews with the manager and heads of specialized departments of the selected small agricultural enterprise. The selected enterprise (its legal form is a limited liability company) has been operating in the Czech market for 15 years and employs 35 people. It is active only in crop production (on 200 ha of land, of which 150 ha are used for growing cereals and 50 ha for orchards). Due to the Covid-19 situation, the directed interviews were conducted mainly online. The resulting data were processed in the form of a case study, using a reference preliminary costing for the acquisition of a drone for its use in the crop production. The data collection, analysis and evaluation were carried out from February to September 2021.

4. Results

The presented research specifically focuses on two elements of the concept of agriculture digitization, namely the use of drone technology and GPS in the production of cereals in the Czech Republic, where they belong to the frequently used representatives of modern technologies [14].

An important and used means for digitization purposes in agriculture is the drone - a small unmanned flying device [6]. It can detect pests, dry areas (e.g., it is used to determine the water stress index of plants) or unseeded areas using conventional photographic images and thermal photos. All this leads to the fact that subsequent spraying, fertilizing and other agricultural interventions can be targeted at the area in question, thus saving the farmer´s time, money, necessary substances and at the same time not unnecessarily burdening areas that do not require intervention. Furthermore, the drone imaging can also identify areas that are at risk of erosion. During harvesting operations (or when mowing meadows), the drone can detect the presence of wildlife in the field so that it can be effectively driven away and not killed [15, 16, 17].

Yield maps (the maps containing the yield in specific parts of the field) represent another possible output of the drone activity. Areas with low yields can be used for other purposes on the basis of the drone mapping (pastures, meadows, bio-belts, field thickets, etc.). The accurate records of land blocks can help the farmer to increase the amount of funding received under subsidies.

The Global Positioning System (GPS) is a technology that brings significant benefits for optimizing and eliminating errors when working in the field. Among other things, it contributes to the saving of fuels, fertilizers or spraying, as well as to the collection of data on machine operation that can be used to implement, for example, anti-erosion measures. Furthermore, when GPS is placed in harvesting machines (e.g., combine harvesters) equipped with special software, it is used to create yield maps. Using these tools, the operator can monitor the running yield of a given field in real time. Other machines, such as tractors, can be similarly equipped to precisely record the amount of seed, fertilizers, or sprays applied to a given area [18].

The small agricultural enterprise surveyed has 200 ha of arable land under cultivation, of which 150 ha are used for cereal cultivation. The enterprise currently owns one modern combine harvester, equipped with GPS, which (thanks to the additional software) is used to create yield maps. Other vehicles owned by the enterprise also have GPS technology in place, but it is currently used only to monitor the fuel consumption. The remaining harvesting machines are of an older production year and so they are not equipped with these modern systems (GPS). The purchase price of a modern combine harvester is approximately between CZK 5 million and CZK 10 million, as are the prices of other necessary modern machines for crop production. Small and medium-sized enterprises, which were focused on in this research, find it more difficult to obtain potential subsidies (primarily due to the high administrative

demands on processing the subsidy applications).

In general, as Karel Dobeš, the Vice Chairman of the Board of Directors of the AMSP CR [12] states, "the use of modern digital technologies places relatively high demands on the design of the overall system and the integration of individual system components. The costs of designing and implementing a complex system usually exceed the possibilities of self-employed persons."

For the majority of small and medium-sized agricultural enterprises or natural persons, the complete replacement of the fleet is not feasible for financial reasons. The use of only one combine harvester with GPS, as implemented by the selected company, proves to be of little benefit. Although the yield maps are of good quality, they are only available for a certain part of the fields actually harvested by this combine harvester. Given the need to harvest grain quickly after its ripening, it is not possible to wait for only one of several machines to do the job. The generated yield maps thus offer only limited possibilities for comparison (e.g., of individual fields) or trend monitoring. The use of GPS technology in harvesters and tractors certainly has its benefits - the creation of yield maps, the better control of the application of seeds, fertilizers and protective agents to individual fields, the possibility of optimizing the movement of machinery in the field, the control of fuel consumption, etc. However, it has a significant disadvantage because the data are collected at the moment of performing the activity. Thus, the selected machine must actually enter the field and perform the activity.

Therefore, the results are obtained only after the specific operations have been performed. In contrast, drone monitoring activities can be carried out repeatedly throughout the year and thus lead to the prevention of certain negative events (drought, pests, weed infestation, etc.). The basic managerial problem when introducing drone digitization into the practice of a small agricultural enterprise is the decision whether it is more appropriate to acquire and operate one´s own drone or to use the already available services of external entities offering "tailor-made" solutions (from the technical, personnel and cost point of view).

The cost of acquiring a drone as part of the agricultural enterprise´s assets

One of the options that the surveyed company is considering in its decision making is purchasing a drone and its operation by the company´s employees. The considered costs are shown in the reference costing in Table 1.

Table 1. Reference costing for the acquisition of a drone

Cost item	Amount (in CZK)
Drone for agricultural purposes (the purchase price)	120,000
Personnel training	13,000
Accessories	12,000
In total	**145,000**

Source: Own processing according to the internal documents of the company, 2021.

Since December 31, 2020, it is necessary for a drone operator in the Czech Republic (both natural and legal persons) to register and pass the qualification competence test online (conducted by the Civil Aviation Authority [19]). In case of the drone flights up to 120 m above the ground, the registration is free of charge, in the case of flights at higher flight levels, the registration is charged at CZK 12,000. The agricultural enterprise surveyed assumes that the first option (up to 120 m above the ground level) will be sufficient. The company also assumes that the drone would be operated by agronomists (as part of their existing job content), i.e., no additional labor costs would be incurred. The estimated service life of the drone is 5 years, and the purchase price also includes the purchase of sufficient number of batteries for several hours of operation. The data processing software for agricultural accounts is

included in the purchase price of the specialized drone.

When choosing a drone, it is necessary to think about the purpose for which is drone acquired - who will use it and in which environment. Some important characteristics of the drone that need to be taken into account during its selection are shown in Table 2.

There are many drones on the market, but not everyone is suitable for use in agriculture. One of the key parameters is the number of motors and the type of wings (rotary-wing, flapping-wing or fixed-wing). Both factors affect flight speed, flight stability and drone carrying capacity. A larger number of motors and fixed wings allow a greater drone load capacity and thus the ability to carry, for example, a spray tank. For larger fields, fixed-wing drones are more recommended (a larger number of sensors are usually placed on them) [20, 21]. This type of drone is shown in Fig. 1.

Fig. 1. Drone ATMOS with fixed-wing for agriculture
Source: [22]

The different type of drone shows on Fig 2. This type is characteristic by its low weight and easy use (so it is easier to transfer it between distributed fields). But to its short time at the air, it is useful only for small fields.

Fig. 2. DJI Mavic Mini Fly Combo.
Source: Own processing, 2021

The drone can have a built-in camera, another possibility is to buy a drone without a camera and then buy camera separately. For agricultural purposes, it is more appropriate to purchase a drone with a built-in camera. The user avoids problems with compatibility between the drone and the camera. The battery life and the range of the remote control must be considered in the context of the size of the fields on which the company farms [20].

Table 2. Some technical parameters of drone selection

Parameters	Meaning
Number of motors	The number of motors affects performance, stability, speed and carrying capacity
Type of wings	Rotary-wing, flapping-wing or fixed-wing) affects stability
Camera	Without camera: Lower drone price, lower weight The camera can be purchased separately according to customer choice
	Built-in camera: Easier to use (drone and camera compatibility guaranteed, easy to use)
Video resolution (camera parameter)	HD to 8K – The quality of the recording may be related to the flight altitude - i.e. the amount of area captured in one flight
Frame rate (FPS – Frame per Second) (camera parameter)	The number of frames affects the amount of data in the video (its smoothness)
Battery	Removable / built-in Determines the amount of time a drone can be in the air
Remote control range	The distance that the drone can work without losing control
Independence & Control	The ability of the drone to fly along a set route or return safely in the case of signal failure, etc. Using a driver with a monitor, using a mobile phone or tablet
Weight	It affects the manipulability of the drone and its agility in the air
Speed	Speed in the air and take-off speed
Control	Using a driver with a monitor, using a mobile phone or tablet

Source: Own processing according to [23], 2021

It is also important to consider the ability to properly analyze the data obtained from the drone and then use results further. The entrepreneur should ensure that the drone and the obtained data are compatible with the used software and also with the ability of employees. Employees should be able to process the data for the responsible workers and thus use all the knowledge in practice.

Like everything, it is necessary to compare the technical parameters with the possibilities (financial and technological) of the farmer. There is no one drone suitable for every farmer. But in the wide range, every farmer can choose according to their needs and possibilities.

The cost of using the comprehensive drone services of a specialized company

The calculation for hiring an external company to operate the drone and to create the outputs has been based on the assumption that external services would be used throughout the year for the total area under cultivation (250 ha) of the agricultural entity. In addition to the drone operation itself, analytical outputs are also included in the estimated costs - that is the processing of the field map, the identification of threatened areas, the map of the expected yield, etc. The total costs are recorded in Table 3.

Table 3. Reference costing for purchasing the drone services for 1 year

Cost item	Amount (in CZK)
Drone lease	42,500
Personnel services	45,000
Complementary services	17,500
In total	**105,000**

Source: Own processing according to the internal documents of the company, 2021.

A survey conducted among companies that offer comprehensive drone services for agricultural purposes has shown that most companies allow payment for services on a pay-as-you-go basis, i.e., the total amount would be paid throughout the year, with one-twelfth of that value paid each month.

In addition to the financial costs, the surveyed agricultural enterprise has also considered other aspects in its decision making. In the long term, it would be worth buying its own drone (with an estimated lifetime of 5 years). However, using the services of an external company (in the Czech Republic there are a number of companies that provide comprehensive services – "drone as a service"), which provides a tailor-made solution, brings a number of indisputable advantages.

One of the reasons is a lack of experience with the handling of drones and working with the relevant software. Although the company does not expect the need to recruit more staff, it cannot be ruled out that operating the drone and the subsequent data processing and analysis will, in practice, require more time than the currently available capacity of the selected staff (agronomists).

Another reason is the necessity to register the agricultural enterprise as a drone operator and the need to monitor other legal regulations, which must be complied with in order to operate a drone. Taking care of the drone - i.e., its maintenance, charging sufficient number of batteries and possible repairs of the drone - may also be the factor that discourages the agricultural enterprise from purchasing and operating the drone on its own. Another issue to be discussed is the use of drones for activities that are not directly offered by external companies (e.g., game scaring from fields and meadows, etc.); therefore, the possibility of including such activities is a question for the further negotiation of the service price with potential suppliers of external services (in the current energy situation in Europe, a further increase in the cost of external services can be expected).

In the current economic conditions, the management of the agricultural enterprise surveyed is inclined towards the more costly, but otherwise more convenient option of purchasing external services directly from specialized companies.

5. Conclusion

The constantly growing human population places more and more demands on the production of agricultural commodities in order to feed the population. Most countries in the world have committed themselves to the principles of the sustainable approach to agriculture. This policy aims to "*maximize the yield of production with minimum environmental impact*" [24, 25]. The use of modern technologies is meant to make the farmers´ work easier and help to increase produce within the same agricultural areas - agricultural land is shrinking because of ever-increasing housing development [26].

The transition to the Agriculture 4.0 concept is a long-term trend within the contemporary society, on which there is also a scientific consensus [1]. The use of technologies may be faced with the lack of funds, the incompatibility of individual methods and tools, the fear of introducing new technologies, and the lack of manpower or its lack of skills [1]. The submitted paper also highlights the costs

associated with the acquisition and operation of new technologies. Although their affordability is increasing as they are becoming more widespread, many issues still remain unresolved. As stated by Lytridis et al. [27] energy requirements limit the number of operations that these tools are capable of performing. And energy intensity is also a key and highly topical issue given the skyrocketing energy prices on world markets (autumn 2021) and the EU´s commitment to the so-called "Green Deal" [28].

As noted by Brant [29] there is not yet a definite answer to the energy benefits of Agriculture 4.0 concept, but there is no doubt about its contribution to reducing the consumption of fuels, fertilizers and pesticides. The application of Agriculture 4.0 principles leads to increasing demands on workers, but also on materials in electronic systems, and there is an increasing pressure to build transmission networks.

The paper focuses on the presentation of the selected technologies (GPS and drones) within the development of Agriculture 4.0 concept, which are often used in crop production in the Czech Republic. Drones are becoming increasingly important, especially due to their versatile use [12] The GPS technology installed in machines for field work provides useful information and data (yield maps, fuel consumption, etc.), but in some respects the data for decision-making and analysis are obtained with a delay. On the other hand, drones provide up-to-date data throughout the year, allowing farmers to continuously respond to the specific needs of a given field (drought, weed infestation, pest infestation, etc.).

However, the existing fleet composition of small and medium-sized agricultural enterprises in the Czech Republic is often obsolete and the complete renewal of fleets is beyond the financial means of these entrepreneurs (also due to the difficulty of obtaining funding from subsidy programs). The use of drones in agriculture brings a greater number of analytical outputs and is also more affordable, both when purchasing the drone itself and when hiring a specialized company that will deliver all activities and outputs to the company as a "turnkey project".

The presented research documents the financial costs associated with the use of drones. It is not possible to clearly determine which option of its acquisition (whether the purchase/ or the lease of comprehensive services) is more advantageous and will help to ensure a long-term perspective. Every farmer has their own individual needs and is in a different environment, and the resources used need to be adapted to these factors. There is no denying that GPS technology and drones are an important part of modern agriculture.

The research presented in this paper is limited. The issues of Agriculture 4.0 concept implementation in practice are multidimensional, very closely related to the national legislation, the history and the development of agriculture, the development of networks and the availability of modern technologies. In order to make the conclusions more generalized, the authors of the research are preparing directed interviews with managers of a selected sample of small and medium-sized agricultural enterprises, and they are tracking the benefits of and barriers to the use of digitization in the longer term.

Acknowledgments

This paper was supported by the University of West Bohemia under the project entitled SGS-2020-026 "Economic and financial transformation in the context of digital society".

References

1. Silveira F., Lermen F. H., and Amaral F. G. An overview of agriculture 4.0 development: Systematic review of descriptions, technologies, barriers, advantages, and disadvantages. Computers and Electronics in Agriculture, vol. 189, 106405 (2021). https://doi.org/10.1016/j.compag.2021.106405
2. Pierce F. J., and Nowak P. Aspects of precision agriculture. Advances in agronomy, vol. 67, pp. 1-85 (1999). https://doi.org/10.1016/S0065-2113(08)60513-1
3. Shannon D. K., Clay D. E., Kitchen N. R., French B. W., and Mathew F. M. Precision Agriculture Basics. Madisown (WI): Crop Science Society of America Soil Science Society of America (2020).
4. Rapela M. A. Fostering Innovation for Agriculture 4.0: A Comprehensive Plant Germplasm System. Cham (CH): Springer Nature Seitzerland AG (2019).
5. Bollini L., Caccamo A., and Martino C. Interfaces of the Agriculture 4.0. In Proceedings of the 15th International Conference on Web Information Systems and Technologies (WEBIST 2019), Wien (2019).
6. Rose D. C., Wheeler R., Lobley M., Winter M., and Chivers C. Agriculture 4.0: Making it work for people, production, and the plane. Land Use Policy, vol. 100, 104933 (2020). https://doi.org/10.1016/j.landusepol.2020.104933
7. Liu Y., Ma X., Shu L., Hancke G. P., and Abu-Mahfouz A. M. From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. IEEE Transactions on Industrial Informatics, vol. 17, pp. 4322-4334 (2020). https://doi.org/10.1109/TII.2020.3003910
8. Zhai Z., Martínez J. F., Beltran V., and Martínez N. L. Decision support systems for agriculture 4.0: Survey and challenges. Computers and Electronics in Agriculture, vol 170, pp. 2-16 (2020). https://doi.org/10.1016/j.compag.2020.105256
9. Kilavuz E. and Erdem I. Agriculture 4.0 applications in the world and transformation of Turkish agriculture. e-Journal of New World Sciences Academy, vol. 14, pp. 133-157 (2019). http://doi.org/10.12739/NWSA.2019.14.4.3C0189
10. Luu T., Tien D. L., Thi N., and Hiep K. The Revolution of Agriculture 4.0 and Sustainable Agriculture Development in Vietnam. In: T. T. Quyen (Ed.), Emerging issues in economics and business in the context of international integration, pp. 317-328 (2017).
11. European Commission Modernising and simplifying the cap. Socio-Economic challenges facing EU agriculture and rural areas. Available online: https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/key_policies/documents/soc_background_final_en.pdf, last accessed 2021/11/02.
12. AMSP ČR. Digitální farma - Srpen 2019 Závěrečná zpráva. Available online: http://amsp.cz/wp-content/uploads/2019/08/Digita%CC%81lni%CC%81-farma_v%C3%BDsledky-kom.-AMSP-%C4%8CR.pdf, last accessed 2021/11/10.
13. Kocsis J., and Major K. A General Overview of Agriculture and Profitability in Agricultural Enterprises in Central Europe. In: Bryła P. (Ed.), Managing Agricultural Enterprises. Palgrave Macmillan, Cham (2018).
14. CzechTrade. Zemědělci jsou v digitalizaci dále než průmysl. Available online: https://www.businessinfo.cz/clanky/zemedelci-jsou-v-digitalizaci-dale-nez-prumysl/, last accessed 2021/11/01.
15. Lawson M. Cost-Effective Uses of Drones in ICT for Agriculture. Available online: https://www.agrilinks.org/post/11-cost-effective-uses-drones-ict-agriculture, last accessed 2021/11/01.
16. Krishna K. R. Agricultural Drones: A Peaceful Pursuit. Apple Academic Press Inc. (2018).
17. European Commission. Drones in agriculture. Available online: https://ati.ec.europa.eu/reports/sectoral-watch/drones-agriculture last accessed 2021/11/09.
18. Suprem A., Mahalik N., and Kim K. A review on application of technology systems, standards and interfaces for agriculture and food sector. Computer Standards and Interfaces, vol. 35, pp. 355-364 (2013). https://doi.org/10.1016/j.csi.2012.09.002
19. Úřad pro civilní letectví. Základní informace ke změnám v regulaci dronů od 31. 12. 2020. Available online: https://www.caa.cz/news/zakladni-informace-ke-zmenam-v-regulaci-dronu-od-31-12-2020/, last accessed 2021/11/01.
20. Precision Agriculture. Choosing the Right Drone. Available online: https://precisionagricultu.re/choosing-the-right-drone/, last accessed 2021/12/10.

21. Kulbacki, M. et al. Survey of Drones for Agriculture Automation from Planting to Harvest. 22nd IEEE International Conference on Intelligent Engineering Systems. INES 2018, Las Palmas de Gran Ccanaria (2018).
22. ATMOS. Available online: https://www.atmosuav.com/, last accessed 2021/12/11
23. Dronpro.cz. Available online: https://dronpro.cz/, last accessed 2021/12/10
24. Abildtrup J. et al. Socio-economic scenerio development for the assessment of climate change impcts on agricultural land use: a pairwise comparison approach. Environmental Science and Policy, vol. 6, no. 1, pp. 101-115 (2006). https://doi.org/10.1016/j.envsci.2005.11.002
25. Dey A., and Mukherjee, S. Importance of Biofertilizers in sustainable Agriculture. Agrobios Newsletter, vol. 18, no. 8, pp. 37-38 (2020).
26. Almadani B., and Mostafa S. M. IIoT Based Multimodal Communication Model for Agriculture and Agro-Industries. IEEE Access. vol. 9, pp. 10070-10088 (2021). https://doi.org/10.1109/ACCESS.2021.3050391
27. Lytridis C., Kaburlasos V. G., Pachidis T., Manios M., Vrochidou E., Kalampokas T., and Chatzistamatis S. An Overview of Cooperative Robotics in Agriculture. Agronomy, vol. 11, 1818 (2021). https://doi.org/10.3390/agronomy11091818
28. Siddi M. The European green deal: Assessing its current state and future implementation. Climate Policy, vol. 16, pp. 543–547 (2020).
29. Zemědělský svaz ČR Hlavním úkolem současné doby je přechod od precizního zemědělství k Zemědělství 4.0! Available online: https://www.tripartita.cz/zemedelsky-svaz-cr-hlavnim-ukolem-soucasne-doby-je-prechod-od-precizniho-zemedelstvi-k-zemedelstvi-4-0/, last accessed 2021/11/07.

Strategic Design for Leather Tannery Industries

Mayra Paucar S.[1[0000-0003-2754-609X]], Pablo Amancha P.[2[0000-0003-1502-6118]], Jorge Santamaría Aguirre[1[0000-0001-7896-3657]], Marcelo Pilamunga Poveda[1[0000-0001-9736-9899]]

[1] Universidad Técnica de Ambato, Ecuador, 180207
[2] Pontificia Universidad Católica del Ecuador Sede Ambato, Ecuador, 180207

https://doi.org/10.33847/2686-8296.3.2_7

Received 01.10.2021/Revised 07.11.2021/Accepted 09.12.2021/Published 28.12.2021

Abstract. The province of Tungurahua in Ecuador has a participation of 76% at the national level within the leather production chain with the production of raw materials, these are used in the manufacture of 80% of footwear in the country. These production processes generate environmental problems, the greatest impact being on the water, due to the dumping of liquid waste with chemical contents in rivers, this generates bad odors, gases, smoke and polluting solid waste from this process. According to data from tanneries, approximately 88% of the raw material is transformed, while the remaining 12% is considered as waste (leather trimmings), which are used in other manufacturing processes (footwear, textile industry, leather goods and others), the chip obtained from the process is compressed and discarded. As a preventive action, a process is proposed to generate new products through the use of waste obtained from leather processing. The design methodology is the Double Diamond: discover, define, develop and deliver a new product that reuses this waste, a block of dimensions 150 x 75 x 5 mm and 300 grams of weight was obtained, which can be applied in the sector of construction and allow the transition from a linear production process to a circular process.

Keywords: strategic design, solid waste, new products, leather industry, leather industry.

1. Introduction

The current production paradigm is mainly based on the linear economy, i.e., "take, make and dispose"; however, the circular economy also considers the recovery phase and some industries are currently implementing these processes. The province of Tungurahua has a large participation within the leather production chain, with 76% [1] in the production of tanned hides at the national level that are used as raw material for the manufacture of 80% of footwear in the country and other uses.

The production processes of tanneries generate environmental problems, with a greater impact on water due to the dumping of liquid waste containing chemicals in streams and rivers, generating bad odors, gases, smoke and solid waste as pollutants from this process. The manufacturing sector in Ecuador has a 13% share of GDP [2] and is made up of 24 subsectors according to the Expanded Classification of Economic Activities (ISIC). Subsector C15 - Manufacture of leather and related products, has as its main economic activities [3]:

- ✓ Production of footwear, boots and similar articles.
- ✓ Manufacture of artificial leather (regenerated), chamois and parchment leathers, patent and metallized leathers.
- ✓ Activities of fleshing, shearing, shearing, depilation, fatliquoring, tanning, bleaching, dyeing, dressing of skins and leathers of fine skins and leathers with hair on.
- ✓ Manufacture of suitcases, backpacks, handbags and other similar articles.

According to the Coordinating Ministry of Production, Employment, and Competitiveness, the textile and leather intermediate and final industries (IIF) generate a greater contribution to this industry, with a 75.6% share in the province of Tungurahua in handicraft, tanning, leather, and footwear activities.

The leather tanning process is important for obtaining raw material for manufacturing products. Consumers perceive leather as a natural material and associate it with sustainability, so the leather industry will adopt sustainable criteria by implementing circular processes, orange economy, and others in its industries. The use of tanned leather is destined for the production of some products worldwide, being used for the manufacture of furniture, accessories, clothing and others. The key areas in leather production are: resources (hides/skins, chemicals, water, energy), emissions generated during the production process (solid, liquid and gaseous waste), quality and attractiveness of the final product (durability) and use of the product after its useful life [4].

Considering that Ecuador processes more than 350,000 hides per year approximately, an estimated 1,500 tons of leather shavings are generated annually and destined for landfills. Several alternatives have been proposed for the valorization of leather shavings, the most used is directly as an adsorbent of: dyes, motor oils, chromium and arsenic present in wastewater; as well as in reinforcement of composite materials with polymeric matrices of rubber, polyvinyl chloride, polyvinyl pyrrolidone and polyvinyl alcohol (PVA) [5].

It has been shown that one of the viable methods for utilization of chromium-containing leather solid waste resources has been to obtain materials for electrostatic flocking, regenerated leather, adsorbents and construction materials through simple treatment steps without dechroming [6].

In Bangladesh they are using the chip from the debinding process through clay stabilization processes to obtain bricks. For sustainable industrial production they have ensured the mechanical properties of bricks as building materials, in addition, their manufacturing process is energy efficient compared to conventional processes. They have manufactured with different percentages of chip composition (10%, 20%, 30% and 40% in dry weight) in controlled laboratories and in the field, has met all the properties required by a building material such as strength, water absorption, shrinkage, weight loss on ignition and bulk density. The bricks with incorporated chips obtained a compressive strength ranging from 10.98 MPa to 29.61 MPa, with water absorption ranging from 7.2% to 20.9%, and met the established ASTM standards for construction materials [7], [8].

The tanneries in Tungurahua do not currently have circular processes that allow for the use of the waste obtained from the leather trimming process; therefore, the final destination is the landfill in the city of Ambato in Chachoán. Only about 50-55% of the collagen actually ends up as finished leather, so tanneries generate large amounts of solid waste. In fact, the utilization and/or disposal of solid waste is today one of the most difficult challenges [4]. The goal of a sustainable strategy is, as much as possible, to produce no waste and to convert any potential waste into an effective and valuable product. Therefore, it is important that the waste does not contain chemicals of concern, they would only limit its further use.

Within the objectives of tanneries is to have lower water consumption, better absorption of chemicals, decrease the use of hazardous and/or banned substances (Substances of Very High Concern, SVHC), improve the quality/reuse of solid waste and reduced content of specific pollutants such as heavy metals and electrolytes contained in them for their processes [4].

Through the application of strategic design can be obtained the planning of new products and services that reduce waste, reuse waste generated in the industry and through strategies minimize negative impacts on sustainability, to provide benefits

for both the economic, social and environmental sector, by extracting the maximum value from raw materials, products and waste; promoting energy savings and reduction of greenhouse gas emissions [9].

To address this problem, using strategic design and user-centered methodologies, problems or opportunities can be discovered to generate proposals for new products that use waste, reincorporating it into the production chain, which allows moving from a linear production model to a circular one, helping to reduce environmental pollution.

2. Literature Review

Waste management and strategic design

Design methodologies emphasize the importance of investigating user needs and understanding the situation in which a product can be generated or improved. Strategic design through the application of criteria is relevant; design is a complex activity, and failure to consider new strategies that propose the creation or improvement of these processes can be costly in terms of time, people and money for companies [10], [11], [12], [13], [14].

The disposal of solid waste helps to achieve sustainability objectives, the reuse of leather shavings produced in the trimming process is an opportunity that can be presented through the application of strategic design, in which the design process can start from an initiative of the company originated in a detected opportunity (solid waste) or as a result of a proposal of the designer that includes a concept of innovation and product that is interesting for the context and the environment, generating a circular system where the result is the obtaining of a strategic object that proposes values, anticipates the needs of the consumer and is environmentally friendly [15], [16].

The amount of solid waste generated will depend on some variants such as the raw material processed (heavy skins, very fatty skins and light skins), differences in the tanning process and substantial variations in the water content in the waste. Commitment to tannery management is a prerequisite for good environmental performance. Technology alone is not enough; it is accompanied by good housekeeping measures. A key to good performance is awareness of the inputs and outputs of the process with respect to material characteristics, quantities and their potential environmental impacts.

Taking into account the criteria that guarantee better environmental performance, as well as the technological criteria that focus the properties of the final product. In addition, pollution will maintain a bearable level with a reduction of spills, accidents, water waste and chemical use. This can be achieved by choosing appropriate techniques, good maintenance and control of the operation, by monitoring and adjusting process parameters, and good training of personnel [4].

Solid waste management is currently one of the most difficult tasks in environmental management: an established method of reuse and/or disposal may become unfeasible in the short term, either due to changes in regulations or for commercial reasons. The image of solid waste handling in a tannery is a reflection of the company's competence, responsibility and social and environmental commitment. The leather tanning industry is considered an economic activity for the province of Tungurahua that contributes directly to the leather and footwear cluster. This process at a general level produces by-products that have contributed to the environmental footprint, as part of the tanning process it generates large volumes of liquid effluents and solid waste, specifically in what has to do with rawhide trimmings, fleshings, keratin waste, polishing powders and post - tanning leather shavings [5].

The discipline of Design in recent years has been transforming and evolving according to the new challenges of the environment and context. In recent decades, designing with the user in mind and involving sustainability in product development (PD) has generated new design currents [17]. Strategic design management can be considered as an integrated network that does not require a certain sequence and needs the interactivity of some important aspects such as [18]:

- ✓ Idea: temporary formulation of the product program.
- ✓ Preliminary analysis regarding the company, market, society, product sustainability, professional management and project feasibility.
- ✓ Strategic objectives of the project: thematic goals and structure of the objectives.
- ✓ Results to be obtained.
- ✓ Project communication.

For the development of new products and design ideas, they will not be accepted or rejected without first being subjected to a verification analysis. In this way, they are considered as temporary and although they can be withdrawn, they can give rise to others, by analogy or contraposition [15]. It is beneficial to discard within the initial idea, the concept and eventually the prefiguration of the object, which may or may not be present. Once the idea has been deepened, it is important to define and evaluate the concept of the product in all the aspects prioritized in the proposal [15]. The prefiguration is part of the design process, it constitutes a pre-design and suggests a materialized solution of the object, in addition, it allows the evaluation of the product proposal.

The sustainability of the product is considered as an unavoidable condition for the social responsibility of both the design and the company, so that there is a concordance between the economic system, the social and the environment, so it will be part of the aspects to be considered in the new project. Although sustainable development may face a conflict of interests due to the need for expansion of the economy of emerging and developed countries and, on the other hand, the aspiration of a large part of society to preserve the environment. The low industrial development of the peripheral regions can be taken as an opportunity to develop policies to control environmental pollution, but on the other hand, the globalization of business, production and marketing inexorably transfers the interests of the central countries to the peripheral ones.

The identification and sizing of trends in inputs, processes, human capacities, impact on the environment, technology and innovation are the challenges that Ecuador faces in meeting the goals. Therefore, it cannot fail to recognize that cultural and social phenomena determine the change in consumption habits and behavior of human beings, i.e. the SDGs are a macro trend in the medium and long term, directly related to economic, social and political factors that will influence lifestyle and consumption decisions.

Macrotrends such as the incorporation of the Sustainable Development Goals, specifically goal 12: Responsible Production and Consumption invites companies in general to the efficient management of shared natural resources and the way they eliminate toxic and polluting waste, industries, businesses and consumers are called to reduce, reuse and recycle waste generated in these processes to achieve sustainable patterns of consumption by 2030, there is less than a decade to achieve this goal, it is important that industries reduce their ecological footprint through changes in production processes and consumption of goods and services [19].

3. Data and Methodology

The R + D + i (research, development and innovation) is a tool that allows the application of sustainable design in a comprehensive manner, where its principles are the harmony between the economic, social and environmental sectors without compromising the needs of future generations, also considering those of the present, it is directly linked to the achievement of the SDGs. Moreover, design has evolved and now allows consumers to interact directly with what they feel, what they want and need. The personalization of products, spaces and services are becoming short-, medium- and long-term trends.

Design Methodology: double diamond

In the field of user experience (UX) for product development, new concepts must be taken into account. Synthesizing and organizing all this new terminology has led to the development of new design methodologies. Concepts such as: Design Thinking, Design Process, Double Diamond Method, Simplex Method of Creative Problem Solving and the different phases of the UX process, have been formed with similar elements. In the different stages, each one has its particularity, but all of them can overlap each other is, the same path, i.e. all design disciplines share the same creative process.

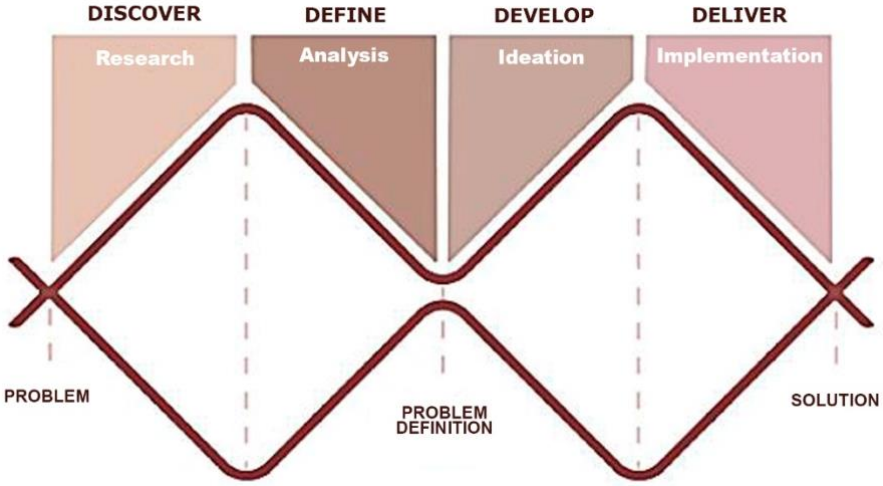


Fig. 1. Stages of the double diamond methodology. The methodological proposal is represented with two diamonds (rhombuses), where the first one symbolizes the Creative process, composed of two stages "Discover and Define", and the second one the Prototyping process, composed of two stages "Develop and Deliver".
Source: Wuth P., Negrete S., Guevara G., & Hojman A. [20]

Each specialization in design has different approaches and ways of working, the Double Diamond methodology is considered as a simple visual map of the design process for the development of a product and / or service [21]. It is based on the initial divergence of thought that investigates several ideas and different points of view and the subsequent convergence of this, as the process progresses, the best findings or solutions that respond to the challenge to be addressed must be selected.

By means of the Strategic Design, the types of waste that are obtained in a tanning process were identified: skin trimmings, raw meat, shavings, leather sanding and sludge. Of these, shavings were selected because of their environmental impact due to their chemical composition (chromium) and because they reduce the volume

of waste in Ambato's landfills. Applying the Double Diamond methodology, a new product has been developed with functional, technical and aesthetic attributes that can be used in the construction sector due to its high growth and innovation in designs and materials.

4. Results

Sustainable design allows the interaction and harmony of the social, economic and environmental sectors, obtaining a new product that meets and satisfies the proposed methodology becomes a challenge. Elaborating the concept of the prototype will comply with specific characteristics such as the reuse of waste from the leather trimming process of tanneries, environmentally friendly, that will boost the local economy, will comply with technical characteristics, functionality and that will be aesthetically attractive to the consumer, makes it possible to become an integral product (insight) that harmonizes each of the needs through the proposed strategic design methodology.

The solid residue coming from leather lowering process has a highly organized structure in the form of fibers ø = 100 nm, which are in parallel and very close together, have a significant protein percentage of 78.64 - 78.00% and trivalent chromium of 3 - 3.3% [22]. The same that according to several toxicity manuals do not represent a potential risk, when inhaled or ingested; however, it is necessary to mention that this residue has a possibility of oxidation from CrIII (chromium 3) to CrVI (chromium 6). Acid rain can leach CrIII-containing waste from landfills causing possible oxidation; another possibility of oxidation is by air over a wide pH range. According to Gibbs free energy law, spontaneous oxidation can occur in an acidic or alkaline medium [22], where they mention the possibility of oxidation of CrIII to CrVI, but they have not performed experimental tests to confirm this hypothesis. In the thermal field, for there to be an oxidation of CrIII to CrVI, the temperature must be greater than or equal to 850°C [23], this data is very important and thermal tests must be considered.

In Ecuador, the construction sector in 2019 was the fifth sector with the highest representation in the country's GDP (8.17%) and had a foreign investment of more than 69 million dollars, generating work for more than 480,000 people - between suitable and unsuitable employment (Ekos CEO Panel, 2020). Bangladesh is currently using the chip from the debinding process through clay stabilization processes to obtain bricks. For sustainable industrial production they have ensured the mechanical properties of bricks as building materials, in addition, their manufacturing process is energy efficient compared to conventional processes [24]. At present, there are materials that have a neutral pH and can be used for these processes.

Fig. 2. Analysis for the development of the new product.
Strategies such as culture, value chain and possible results to be expected are analyzed. In addition, sketches were made in order to conceive the idea in an adequate and concise manner, and later study prototypes to verify the feasibility of the design.

We obtain the block built with stabilizing material, which allows to confine inside it the shavings from the leather deburring process so that the residue does not come out on the surface. The characteristics are determined and fixed by the composition of the block. With the selected option, Chrome content tests will be carried out. For the application of the prototype to confine the chips inside the block and by the physical-chemical characteristics can propose its use for indoor and outdoor spaces.

Fig. 3. Technical data sheet of the proposal. The characteristics of the final prototype are described in relation to cost, composition, weight, dimensions and amount of CrIII.

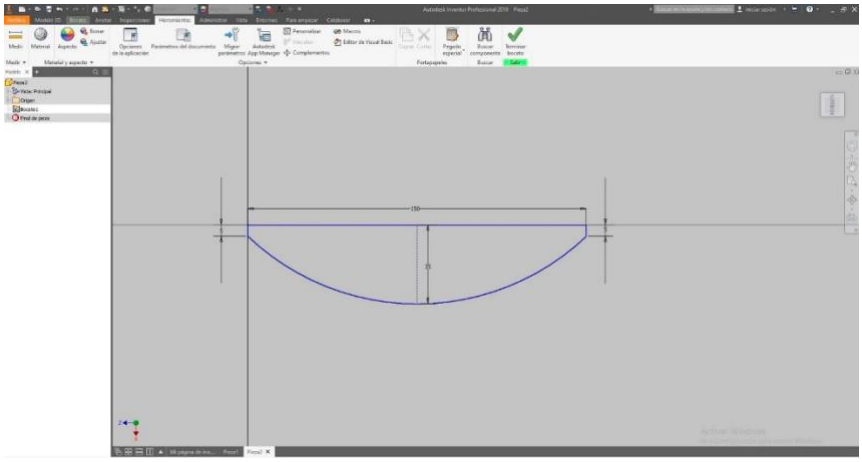


Fig. 4. (a) Computer model of the block. Modeling of the selected morphology with the aid of CAD software (2D).

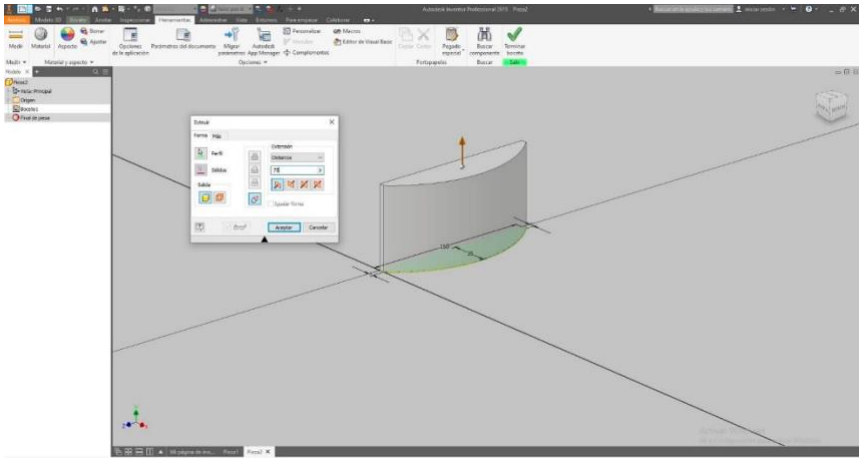


Fig. 4. (b) Computer model of the block. Extrusion process of the selected morphology (3D).

Fig. 5. Rendering of the prototypes of the block with materials

Fig. 6. (a) Proposed application of the block. Application of the prototype in the interior design of shop windows.

Fig. 7. (b) Proposed application of the block. Application of the prototype in the Interior Design of houses.

Product stacking tests were carried out and the results were sent to the ENVIRONOLAB laboratory to test the chromium content in the final product. The Cr values obtained for the long chip samples were 0.74 mg/kg. The maximum permissible limit for heavy metal extraction on a dry basis (acid digestion) for total chromium is 3000 mg/kg. The Double Diamond process can have variations and depends on the role or project. There are two possible uses for the obtained blocks, the first one is the application in commercial display windows and the other one is for applications in living or waiting rooms, as an option for professionals in Interior Design.

With the development of this new product, it is estimated to avoid approximately 120 tons per year of shavings from the leather trimming process of a single tannery. It should be considered that annually in Ecuador this process is estimated to produce more than 1,500 tons of this waste, which are deposited in dumps and landfills of the cities, causing leachates that cause damage to the environment.

5. Conclusion

The theoretical foundation allows us to affirm that Strategic Design relates concepts and languages of design and companies, to achieve through an interactive action the realization of projects, where they define the priority objectives and available resources to respond to the new contexts and problems they present. The generation of waste (solid, liquid and gaseous) in tanneries generates environmental impacts that have not been mitigated. A Strategic Design model allows analyzing and proposing circular processes within tanneries to obtain new products from the use of these wastes.

By means of the Strategic Design model, the types of waste that are obtained in a tanning process were identified: skin trimmings, raw meat, shavings, leather sanding and sludge. Of these, shavings should be selected because of their environmental impact due to their chemical composition (chromium) and because they reduce the volume of waste in the city's landfills. This waste is the raw material for generating new products, reinserting these wastes back into the production chain and preventing them from ending up in the city's landfills, thus reducing the environmental impacts that they can generate when discarded.

A new product has been developed with functional, technical and aesthetic attributes that can be used in the construction sector due to its high growth and innovation in designs and materials. A product was obtained that confined the chip in its interior with a chromium content of 0.74 mg/kg, weight of 300 grams, modular and aesthetically attractive to the user. The weighted cost is 43.29 USD which makes it competitive in the market with respect to the price offered for 3d panels used for the design of spaces that range from 30 to 60 USD/m2.

The Strategic Design Model developed is a contribution to the macro trend of achieving the SDGs, specifically goal 12, in which through the convergent and divergent phases have identified a problem, defined the solution, developed and delivered a new product that responds to the management of waste (shavings) generated in the tanning industry to mitigate the negative effects on the environment.

The application of the Double Diamond Methodology allows the interaction of actors that form a network between academia (researchers), decentralized autonomous governments, industry and the construction sector, who are the appropriate ones to jointly develop research projects that contribute positively to society in general with social responsibility.

6. Recommendations and Future research lines

It is recommended that mechanical strength and water absorption tests be performed on the new product to determine if the confinement of the chip inside the product does not allow chromium leaching under aqueous conditions. In addition, a hydrolysis process can be added for the chip, which will allow the total removal of CrIII.

The Double Diamond Methodology can be used for the generation of new products and services that allow diversifying its use and help the reuse of waste from the leather trimming process. In addition, further research is suggested to determine whether the block can contribute as an acoustic or thermal insulator in its final use.

At the governmental level, incentives can be developed through public policies for companies that reuse their waste to obtain new products, as is done in other countries around the world.

Acknowledgments

Special thanks to the Universidad Técnica de Ambato and jointly to the Directorate of Research and Development (DIDE) that have supported the project "Strategies for the reuse of remnants of the Leather and Footwear cluster of Tungurahua as a contribution to the sustainability of the productive sector", with Resolution No. UTA-CONIN-2020-0300-A.

References

1. Ruiz M., Mayorga C., Mantilla L. y López P., «Gestión económica ambiental del sector curtiembre de Ambato,» *Augusto Guzzo Revista Acadêmica,* vol. 1, nº 17, p. 133, 2016.

2. Banco Central del Ecuador, «Indicadores Económicos,» 2020. [En línea]. Available: https://www.bce.fin.ec/index.php/informacioneconomica.

3. Camino S., Vera S., Bravo D. y Herrera D., «Estudios sectoriales: manufacturas,» Superintendencia de compañías, valores y seguros del Ecuador, Guayaquil - Ecuador, 2017.

4. United Nations Industrial Developmet Organization, «The framework for sustainable leather manufacture,» UNIDO, Vienna - Austria, 2019.

5. Viteri A., Valle V., Bonilla O. y Quiroz F., «Valorización de virutas de cuero post - curtición para la obtención de mezclas termoplásticas,» *Revista Iberoamericana de Polímeros y Materiales,* vol. 17, nº 5, pp. 248 - 260, 2017.

6. Li Y., Guo R., Lu W. y Zhu D., «Research progress on resource utilization of leather solid waste,» *Journal of Leather Science and Engineering,* vol. 1, nº 6, pp. 1 - 17, 2019.

7. M. A. I. Juel, A. Mizan y T. Ahmed, «Sustainable use of tannery sludge in brick manufacturing in Bangladesh,» *Waste Management, 60,* pp. 259-269, 2017.

8. B. Saha y F. A. B. Azam, «Probable Ways of Tannery's Solid and Liquid Waste Management in Bangladesh–An Overview,» *Textile & Leather Review, 4(2),* pp. 76-95, 2021.

9. S. I. Hallstedt, «Sustainability criteria and sustainability compliance index for decision support in product development,» *Journal of Cleaner Production,* pp. 1-16, 2015.

10. Byggeth S., Broma G. y &. R. Karl-Henrik, «A method for sustainable product development based on a modular system of guiding questions,» *Journal of Cleaner Production,* vol. 15, nº 1, pp. 1-11, 2007.

11. Akturk M., Serkan, Abbey J. y Geismar H., «Strategic design of multiple lifecycle products for remanufacturing operations,» *IISE Transactions,* pp. 1-28, 2017.

12. Baez I., Carrillo C., Castelblanco O., Betancourt F., Leguizamón G., García R. y Mendoza D., «Product Design Methodology under the structure of Innovation and Creativity. Review study,» *ESPACIOS,* vol. 39, nº 11, pp. 20 - 42, 2018.

13. N. d. P. B. v. d. G. B. Bocken, «Product design and business model strategies for a circular economy,» *Journal of Industrial and Production Engineering,* vol. 33:5, pp. 308-320, 2016.

14. Ramírez R. y Ariza R., Diseño de productos : una oportunidad para innovar : programa: gestión del diseño como factor de innovación, Buenos Aires: INTI, 2015.

15. R. Leiro, Diseño: estrategia y gestión (1a ed.), Buenos Aires: Ediciones Infinito, 2008.

16. S. M. A. A. Mariam A., «A simulation-optimization model for sustainable product design and efficient end-of-life management based on individual producer responsibility,» *Resources, Conservation & Recycling,* vol. 140, pp. 246-258, 2019.

17. Royalty A., Ladenheim M. y Roth B., «Assessing the Development of Design Thinking: From Training to Organizational Application,» *Understanding Innovation,* 2015.

18. S. De Pietro y P. Hamra, Diseñar hoy: visión y gestión estratégica del diseño, Buenos Aires: Nobuko, 2010.

19. Comisión Económica para América Latina y el Caribe , «La Agenda 2030 para el Desarrollo Sostenible en el nuevo contexto mundial y regional: escenarios y proyecciones en la presente crisis,» LC/PUB.2020/5, Santiago, 2020.

20. Wuth P., Negrete S., Guevara G. y Hojman A., «DI: Diseño para Innovar,» Pontificia Universidad Católica de Chile, Santiago de Chile, 2019.

21. O. Castillo, «Designthinking y el Método del Doble Diamante para el desarrollo de prototipos de Emprendimientos o StartUps,» *Perspectivas: Revista Científica de la Universidad de Belgrano, 2(2),* pp. 84 - 91, 2019.

22. H. Lakrafi, S. Tahiri, A. Albizane y M. El Otmani, «Effect of wet blue chrome shaving and buffing dust of leather industry on the thermal conductivity of cement and plaster based materials,» *Construction and Building Materias,* pp. 590 - 596, 2012.

23. X. Elias, Reciclaje de residuos sólidos industriales: residuos sólidos urbanos y fangos de depuradora (2 da edición ed.), España: Díaz de Santos, 2009.

24. I. Ariful, M. Al y A. Tanvir, «Sustainable use of tannery sludge in brick manufacturing in Bangladesh,» *Waste Management, 60,* pp. 259 - 269, 2017.

Aims and Objectives

Published online by ICS two times a year, Journal of Digital Science (JDS) is an international peer-reviewed journal which aims at the latest ideas, innovations, trends, experiences and concerns in the field of digital science covering all areas of the scholarly literature of the sciences, social sciences. The main topics currently covered include: Artificial Intelligence Research; Digital Economics, Education, Engineering, Finance, Health Care.

The main goal of the journal is the effective dissemination of original incites/results generated by the human brain and presented/reflected in articles using modern information/digital technology.

Editorial Board

Contact information

Website: https://ics.events

Email: conf@ics.events