



HAL
open science

Enhancing Decision-Making in Product Development: Predicting a medicine-based treatment for a new disease using a Multidimensional Neural Network

Marie Saade, Maroun Jneid, Imad Saleh

► **To cite this version:**

Marie Saade, Maroun Jneid, Imad Saleh. Enhancing Decision-Making in Product Development: Predicting a medicine-based treatment for a new disease using a Multidimensional Neural Network. 37th IBIMA conference, May 2021, Cordoba, Spain. hal-03526959

HAL Id: hal-03526959

<https://hal.science/hal-03526959>

Submitted on 31 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhancing Decision-Making in Product Development: Predicting a medicine-based treatment for a new disease using a Multidimensional Neural Network

Marie SAADE^{1,*} [0000-0003-0520-585X], Maroun JNEID², Imad SALEH¹

¹ Laboratoire Paragraphe (EA 349), Université Paris 8 Vincennes-Saint-Denis, Saint-Denis, France
marie.karam@etud.univ-paris8.fr, imad.saleh@univ-paris8.fr

² TICKET Lab., Antonine University, Hadat-Baabda, Lebanon
maroun.jneid@ua.edu.lb

* Corresponding author

E-mail: marie.karam@etud.univ-paris8.fr (MS)

Abstract. As the COVID-19 pandemic outbreaks all over the world, threatening the lives of millions of people, medical scientists are conducting studies for possible treatments, while several clinical medical centers are currently testing numerous medicines in clinical trials. However, there is no powerful and effective medication recommended so far, in addition to the possibility of clinical trials failure in the context of unbeneficial investments and undesired outcomes.

To address this gap, a new quantitative approach has been proposed in the current paper to monitor the medicine-based treatment development. Precisely, the cure and mortality rates of a potential treatment can be predicted systematically by using convenient predictive multidimensional neural network techniques and based on the quantitative analysis of Big Data for different indicators. More specifically, the diseases' symptoms, along with the corresponding medicines, the active ingredients that compose these medicines, the related patents and publications data, the related reports in clinical trials, etc. can be all considered as effective key performance indicators in evaluating medicine-based product success. This paper covers two main steps: listing all possible combinations of several medicines' ingredients as suggested treatments, then forecasting the cure and mortality rates of each proposed treatment based on the aforementioned dimensions. The present methodology is applied using three types of Multidimensional Neural Networks for comparison purposes: "Wide model" (WNN), "Deep model" (DNN) and "Wide and Deep model" (WDNN).

The findings show that the WDNN model achieves higher prediction accuracy and outperforms WNN and DNN, with a significant prediction accuracy equal to 99.9%. Consequently, addressing this estimation represents a decision support for pharmaceutical and clinical medical centers and a crucial prerequisite step before proceeding with investments decisions in clinical trials and medicines production.

Keywords: Innovative product development. Forecasting medicine-based product success. Multidimensional Neural Networks. Decision support.

1. Introduction

As COVID-19 pandemic outbreaks all over the world, threatening the life of millions of people, medical scientists are conducting studies for possible new therapies, and several laboratories around the world are currently testing and investigating multiple medicines in clinical trials to treat patients from COVID-19. For instance, in May 2020 (National Institutes of Health, 2020), the National Institutes of Health started clinical trials on the Hydroxychloroquine drug combined with the antibacterial drug Azithromycin. In addition, several clinical trials tested ancient systems of alternative medicine such as the traditional Chinese medicine for the treatment of COVID-19 (Xiong, et al., 2020).

Nonetheless, there is no powerful and effective medication recommended so far for treating COVID-19 (Sun, Jiang, Wang, & Liu, 2020), in addition to the possibility of clinical trials failure in the context of unbeneficial and loss-making investments and undesired outcomes (Fogel, 2018). Accordingly, exploring and estimating rapidly and urgently the most appropriate medicine-based treatment for any new pandemic is a crucial need and requires prerequisite steps before proceeding with investments decisions in clinical trials and medicines productions.

The main problematic of this study is to examine how investors or organizations can measure and estimate risks quantitatively at an early stage before investing in a specific product. Hence and more precisely, the research question of the present paper is: how can the success rate of an existing or new medicine-based product be forecasted systematically in order to be a potential promising treatment for a new disease?

Several previous studies proposed different methodologies to identify a potential effective treatment for COVID-19. For instance, a systematic review and meta-analysis of the usage of Chinese herbal drugs have been elaborated as a co-therapy for treating COVID-19, based on a bibliometric analysis and several literature databases such as: ScienceDirect, Google Scholar, China National Knowledge Infrastructure, Embase etc. (Fan, Gu, & Alemi, 2020) (Sun, Sun, & Li, 2020) (Yang, et al., 2020). In addition, multiple network pharmacology tools have been used to predict and analyze the effectiveness and the mechanism of traditional Chinese medicines in the COVID-19 treatment, by mapping the main active compounds of Chinese herbal drugs to the COVID-19 disease-associated genes that represent the viral targets of the drug components (Xiong, et al., 2020) (Sun, Jiang, Wang, & Liu, 2020). Furthermore, a novel approach has been proposed to assess the efficiency of the Immunomodulatory drugs for managing COVID-19 based on two dimensions: a systematic literature search and related reports in 'Clinicaltrials.gov' (Razmi, et al., 2020).

However, and based on the aforementioned listed studies, ancient systems of alternative medicine are considered as a co-therapy and not as a main therapy. In addition, the (Razmi, et al., 2020) study along with other studies focus on a single specific category of drugs with a limited number of key performance indicators or knowledge dimensions. Therefore, each of these approaches has limitations and still needs further investigation and monitoring, reflecting that there is no consensus on the most effective and optimal medicine from COVID-19 so far.

In this regard, to address this gap, the medicine-based product development process can be monitored and evaluated. More precisely, the cure and mortality rates of a potential treatment can be predicted rapidly and systematically by using convenient predictive multidimensional neural network techniques and based on the quantitative analysis of Big Data for different indicators. More specifically, the symptoms of several diseases, along with all types of medicines (antiviral medicines, antibacterial drugs, anticancer drugs, etc.) that have been historically attempted in the treatment of different diseases, the active ingredients¹ that compose these medicines, the patents and publications data related to both the disease and the corresponding medicines ingredients, the related reports in clinical trials including the cure and mortality rates of the attempted treatment, etc. can be all considered as major key performance indicators and effective knowledge dimensions in evaluating medicine-based product success.

This paper covers two main steps: listing all possible combinations of several medicines' ingredients as suggested single-drug or multiple-drugs treatments, then forecasting the cure and mortality rates of each proposed or suggested treatment based on the aforementioned dimensions. The present methodology is

¹ Active ingredients refers to drugs molecules that have a direct biological effect for a disease treatment. (World Health Organization, 2011)

applied using three different types of Multidimensional Neural Network models for comparison purposes: “Wide model” (WNN), “Deep model” (DNN) and “Wide and Deep model” (WDNN).

Accordingly, the Neural Networks are designed to forecast rapidly the optimal effective treatment for a new disease, which can be evaluated and tested by medical centers at later stages in order to validate if the predicted medicines can be considered as a promising and potential treatment.

This systematic prediction system is considered as a crucial need and a prerequisite step before proceeding with investments decisions in clinical trials and medicines production. Therefore, addressing this estimation represents a decision support for medical and pharmaceutical centers in medicine-based product development for treatment of a new disease, and that in order to narrow down their studies areas within a pandemic or a new disease.

The remainder of this paper is structured as following: Section 2 presents the related work on products success forecasting methods. Section 3 presents the research design and the proposed methodology. The experimental setting of the present study and the results are described in section 4. Section 5 concludes the work and points to the study limitations and to future research directions.

2. Literature Review

Independently from the related business type, estimating new products success and identifying high-risk investments before investing in new products is a fundamental necessity, given that they allow organizations to enhance the product development and to understand and estimate the risk of a portfolio based on its return (Chow, 1995).

Precisely, designing an effective predictive analytics system including all process stages, such as data mining, data modeling, statistics, machine learning and artificial intelligence, enables organizations to analyze historical data and then predict new opportunities in order to foster future product development.

For instance, Bhatnagar, Majhi and Sahadev (Sahadev, 2020) suggested a quantitative approach to predict potential successful new products of firms from manufacturing sector industries, using statistical multiple linear regression and artificial neural network techniques for comparison purposes. In addition, a Wide and Deep Neural Network and a Recurrent Neural Network have been implemented by Saade, Jneid and Saleh (Saade, Jneid, & Saleh, Predicting technology success based on patent data, using a wide and deep neural network and a recurrent neural network, 2019) in order to predict a technology success based on patents data. Moreover, Saade, Jneid and Saleh have elaborated another study that aims to enhance decision-making in new product development based on three dimensions: patents, publications and revenues data of several trending technologies, using Multidimensional Neural Networks (Saade, Jneid, & Saleh, 2020).

In the medical sector, a novel approach has been proposed to assess the therapeutic value of the Immunomodulatory drugs for managing COVID-19 based on a two dimensions: a systematic literature that was conducted in several databases to search for available evidence on the efficiency of the Immunomodulatory agents, in addition to related reports in ‘Clinicaltrials.gov’ that was used to extract main findings and output results (Razmi, et al., 2020). An early prediction of the success of the Methotrexate medication for ectopic pregnancy cases has been elaborated using a Retrospective Cohort study and based on the comparison of maternal and gestation characteristics (Levin, et al., 2019). In addition, multiple network pharmacology tools, such as Proteins to Protein Interaction (PPI), Reactome, TCMATCOV platform, etc. have been used to predict and analyze the efficiency of the Shufeng Jiedu (SFJD) capsule, a Chinese herbal drug, in the COVID-19 patients’ treatment. The Comparative Toxicogenomics Database

(CTD) is employed to map the main active compounds of the Shufeng Jiedu (SFJD) capsule to the COVID-19 disease-associated genes which represent the viral targets of the drug components. The latter is obtained from the Encyclopedia of Traditional Chinese Medicine (ETCM), the Bioinformatics Analysis Tool for Molecular Mechanism of Traditional Chinese Medicine (BATMAN-TCM), the Chinese Medicine System Pharmacology Analysis Platform (TCMSP) and related literature records. Disease-associated genes are extracted from the Human Online Mendelian Inheritance Platform (OMIM), the Human Gene Database (GeneCards) and the DisGeNET database (Xiong, et al., 2020).

Furthermore, for the food industry, an artificial intelligence system for predicting success of new product development has been designed in order to identify a proper market-product strategy (Gholamreza & Alireza, 2018). A comparison study has been implemented as well to predict the success of new product development in the clothing industry, using Radial Basis Function and Multilayer Perceptron Neural Network techniques for comparison purposes (Golamreza, Pooya, Kazemi, & Naji-Azimi, 2017). Moreover, a decision-support system for forecasting new product success has been proposed in the (Ching-Chin, Leng, Ling-Ling, & Ling-Chieh, 2010) paper, including four modules: collecting and analyzing data, listing different model templates, choosing the most appropriate model based on data availability and adjusting and calculating the actual forecasts.

Although several previous studies propose to identify future product development, none of them implement Neural Networks predictive tools for forecasting the medicine-based treatment success rate, and using several key performance indicators, such as: the symptoms of several diseases, the medicines that have been historically attempted in the treatment of these diseases, the active ingredients that compose these medicines, the patents and publications data related to both the disease and the corresponding medicines ingredients, several criteria in clinical trials reports, etc.

Accordingly, and based on our research, the current study has not been applied so far when evaluating and monitoring medicine-based product development in this specific context.

3. Research Design

The analysis of Big Data of the aforementioned indicators or dimensions is considered as a quantitative approach, using convenient predictive multidimensional neural network techniques. Precisely, this study investigates the relationship of these dimensions to assess their impact on the medicine-based product development.

The process of this approach is to first prepare the training and the testing datasets of the mentioned dimensions in order to train and validate three types of Multidimensional Neural Network Regression models for comparison purposes: “Wide model” (WNN), “Deep model” (DNN) and “Wide and Deep model” (WDNN).

On one hand, these neural networks are trained and validated based on the following input and output dimensions:

Table 1: Neural Networks input/output dimensions for training and validation.

Input Dimensions		
Dimension 1	The main symptoms of the disease.	Represented as a binary vector containing all distinct symptoms of all diseases: the

		matching symptoms for a specific disease are equal to 1.
Dimension 2	The active ingredients that form the medicines that have been attempted in clinical trials for each disease.	Represented as a binary vector containing all distinct ingredients of all medicines: the matching ingredients for a specific medicine are equal to 1.
Dimension 3	The number of patents related to both the diseases and the related ingredients.	Represented as numeric values.
Dimension 4	The number of publications related to both the diseases and the related ingredients.	Represented as numeric values.
Dimension 5	Participants several criteria used in clinical trials, such as: <ul style="list-style-type: none"> - Age range - Weight category (underweight, normal, overweight) 	Represented as categorical values.
<i>Output Dimensions</i>		
Dimension 6	The cure and mortality rates of the medicines in clinical trials.	Represented as numeric values.

On the other hand, for the prediction phase, the aforementioned dimensions are applied for the candidate diseases, as following:

Table 2: Neural Networks input/output dimensions for prediction.

<i>Input Dimensions</i>		
Dimension 1	The main symptoms of the candidate disease.	
Dimension 2	All possible combinations of several medicines' ingredients, to be considered as suggested treatments. Noting that in the current study, two active ingredients of different medicines are combined at most.	
Dimension 3	The total number of patents related to both the candidate disease and the combined ingredients.	
Dimension 4	The total number of publications related to both the candidate disease and the combined ingredients.	
Dimension 5	The Age range and the Weight category of the group of patients in question.	
<i>Output Dimension</i>		
Dimension 6	It represents the predicted cure and mortality rates. Therefore, the highest cure rates or probabilities of the suggested medicines can be evaluated and tested in clinical trials at later stages to be a promising treatment from such pandemic or disease.	

Why use 'clinicaltrials.gov': Unlike the conventional and homogeneous drug therapy that considers wide patients population, a Personalized Medicine (PM) represents a fundamental management strategy for health care practitioners, targeting specific patient groups (Vogenberg, Barash, & Pursel, 2010). Therefore, the clinical trials can be used to extract data and information related to the disease and the medicines in

question, such as clinical characteristics of a specific type of population (such as age range, weight condition, etc.) and the main findings and output results (Razmi, et al., 2020).

Why use symptoms data: Based on the Personalized Medicine (PM) context and given the fact that each patient may react differently to a same disease, the symptoms of a specific patients group may differ as well, hence the necessity to tune the treatment approach. In addition, the symptoms list can be used in the proposed methodology to match similarity with other diseases.

Why use Patents data: Generally, patents data can be used as key performance indicator in enhancing decision-making in new product development (Saade, Jneid, & Saleh, Predicting technology success based on patent data, using a wide and deep neural network and a recurrent neural network, 2019) (Saade, Jneid, & Saleh, 2020). In addition, research and development on diseases, specifically the number of patents or patents distribution, might offer insights which would be useful to the development of a treatment (Liu, et al., 2020).

Why use Publications data: Similar to the patents dimension, the systematic literature can be conducted to search the available evidence on the efficiency of a medicine in the related disease treatment (Razmi, et al., 2020) (Liu, et al., 2020).

Why use a Neural Network model: Generally, in order to generate higher predictions reliability and accuracy, a predictive model should be trained with Big Data. However, Big Data analytics challenges and issues can be related to data processing, data representation, the fast information retrieval, etc. For this reason, and since a Neural Network model can handle complex large-scale data, it can be implemented to support such tasks, integrating heterogeneous input data (Elaraby, Elmogy, & Barakat, 2016) (Najafabadi M. , et al., 2015). For instance, Neural Networks can outperform traditional models, such as Moving Average (MA), Autoregressive Integrated Moving Average (Adebiyi, Adewumi, & Ayo, 2014), Linear Regression (Gosasang, Chandraprakaikul, & Kiattisin, 2011), in the context of obtaining higher accuracy. Furthermore, given that the current study model will eventually be targeting organizations in a future work, a predictive neural network can be trained and tested using historical Big Data for a dynamic (ALSHEIKH, LIN, NIYATO, & TAN, 2014) and unsupervised learning (Schmidhuber, 2015).

Why use a Deep Neural Network model: Deep Learning model can generalize the relationships in the data, by extracting new or rare combinations by transitivity of correlations (Cheng, et al., 2016).

Why use a Wide Neural Network model: A wide model can learn the exception rules and memorize features interactions or correlations in the historical data (Cheng, et al., 2016).

Why use a Wide and Deep Neural Network model: Although a Deep Neural Network can generalize the extracted representation, it may over-generalize and extract less relevant features (Cheng, et al., 2016). Accordingly, and as a hypothesis, combining the wide model with the deep model can enhance the neural network forecasting by memorizing features correlations and the exception rules.

3.1. Proposed methodology

The current methodology covers different principal objectives explained in Table 1. Noting that these objectives and tasks are applied for all neural network types.

Table 3: Methodology design objectives.

Objective: Data Collection	
1	Listing several historical diseases/pandemics and their related keywords.
2	Listing the main symptoms of each disease.

3	Listing active pharmaceutical ingredients of the medicines that have been attempted in clinical trials for each disease.
4	Extracting the needed information from ‘clinicaltrials.gov’, such as participants’ age range and weight, and their derived values like the cure and mortality rates.
Objective: Database integration	
5	Inserting the collected data into a new integrated database.
Objective: Data Collection	
6	Extracting patents and publications data related to both the disease and the corresponding medicines, based on keywords matching (diseases keywords, medicines terminologies and ingredients keywords).
Objective: Database integration	
7	Manipulating and inserting the collected data into the integrated database.
Objective: Data manipulation	
8	Computing the total number of patents / publications per disease per treatment. Noting that a treatment for a disease can include several medicines, and a medicine can include several active ingredients.
9	Generating all possible combinations of all medicines’ active ingredients, to represent suggested treatments.
Objective: Datasets creation	
10	Grouping the data into training and testing datasets.
Objective: Neural Network implementation	
11	Training the Neural Networks. Noting that the output data represents the cure and mortality rates of the corresponding treatment for each disease.
12	Testing the Neural Networks.
13	Predicting the cure and mortality rates of the suggested treatments for the candidate disease.
Objective: Database integration	
14	Inserting the output data of the Neural Networks into the integrated database.
Objective: Results visualization and Decision-making	
15	Generating the top ten medicines with highest cure rates.
16	Evaluating and testing the probable treatments in clinical trials at later stages.

3.2. Neural Network Structures

The Neural Networks are implemented using TensorFlow² under Python and designed as multidimensional and regression predictive models. The term “regression” refers to the prediction of continuous variables (Bishop, 2006), in the present paper, the estimation of the cure and mortality probabilities or rates.

² The Tensorflow software library is an open-source API and is used for Machine Learning.

Furthermore, the Root Mean Square Prop (RMSProp) optimizer is used for the three Neural Networks in order to speed up the gradient descent (Ruder, 2016).

The Neural Networks consist of the following layers:

Wide model structure:

- **Input layer:** the input nodes represent buckets of data: the input dimensions represent categories of values, noting that numeric dimensions such as the number of patents and the number of publications are also classified in groups.
- **Output layer:** represents the data to be forecasted.

Deep model structure:

- **Input layer:** The input nodes are composed of continuous numbers, where even the categorical values such as patients' age ranges and weight classes are represented as numeric values. For example, for the weight classes, 1 refers to the "Underweight" category, 2 refers to "Normal" and 3 to the "Overweight" category.
- **Hidden Layers:** These intermediary layers make the neural network more adaptable. It includes a specific number h of hidden layers, and each intermediary layer consists of a specific number no of nodes: the first hidden layer contains no_1 nodes, the second hidden layer contains no_2 nodes, etc. and layer h contains no_h nodes.

Noting that these layers could be adjustable based on the loss results and accuracy. The more we have hidden layers, the more the Neural Network output is accurate. Nevertheless, the Neural Network computation will become slower when the number of hidden nodes increase. For this reason, the number of hidden layers and nodes can be adjustable and optimized using specialized methods, such as Competitive Learning Approach that allows to indicate automatically and initially the number of hidden nodes during the training of the artificial neural network (Foram Panchal, 2015).

- **Output layer:** represents the data to be forecasted.

Wide and Deep model structure: The WDNN combines the above structures. Noting that the main categorical dimensions (Dimension 5) represent the inputs of the Wide part, and the numeric and vectors dimensions (Dimension 1, 2, 3 and 4) represent the inputs of the Deep part.

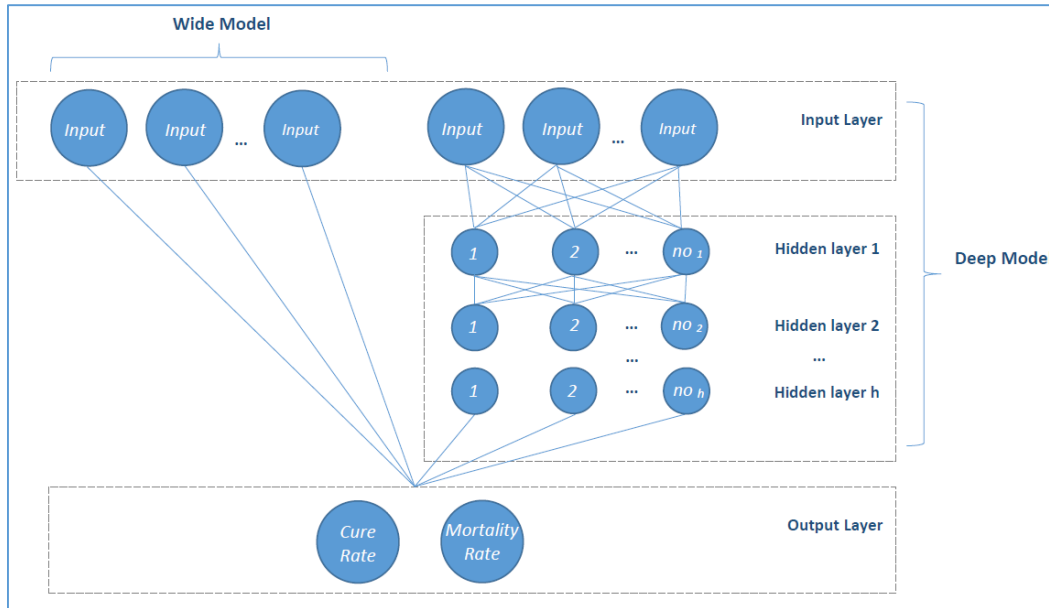


Figure 1: Multi-dimensional Wide and Deep Neural Network.

4. Experimentation

The proposed methodology has been conducted on 13 diseases and 28 active ingredients in order to train the neural networks, then applied on two additional candidate diseases, “Multiple Myeloma” and “COVID-19”, for the testing and prediction phases respectively. Noting that the neural networks can be trained using historical Big Data in a future work, since this predictive model will be targeting actual organizations.

4.1 Data Collection

Diseases and related Symptoms Listing. Different web sources such as World Health Organization, Centers for Disease Control and Prevention, etc. are used to manually extract several diseases and their related keywords and main symptoms.

Medicines’ Active Ingredients Source. A CSV file containing all medicines’ active ingredients is downloaded from the U.S. Food & Drug website.

Clinical Trials Data Source. Several information have been manually extracted from ‘*ClinicalTrials.gov*’ for each disease, such as:

- The medicines that have been attempted in a clinical trial for a specific disease.
- Age range and Weight conditions.
- Number of participants.
- Study results:
 - o ‘Completed’: represents the number of cases who ended normally, and that participants are no longer being treated.
 - o ‘Death’.

Noting that the cure rate and mortality rate are calculated based on the number of participants and the study results: $Cure\ rate = (Completed * 100) / Number\ of\ participants$ and $Mortality\ rate = (Death * 100) / Number\ of\ participants$.

Patent Data Source. Patents identification numbers are extracted automatically using the United States Patent and Trademark Office API (USPTO) which is considered among the richest intellectual property databases, as it issued approximately 10 million patents so far (USPTO, 2018).

Specifically, a script has been developed under Python to search the keywords of each disease and the related ingredients, in the abstracts of all granted patents applications issued before 15 January 2021.

Publications Data Source. The numbers of publications are extracted automatically using the Springer API that includes a wide range of bibliometric resources. A script has been developed under Python to search jointly the keywords of each disease and the related ingredients, in the titles of all publications issued before 17 January 2021.

4.2 Training, Testing and Prediction Datasets Creation

The collected data has been split manually into training and validation datasets, since the amount of data is limited in volume. In addition, in order to visualize the accuracy of the Neural Networks and to compare the actual values with the predicted values, the testing or validation dataset is considered as the prediction dataset, where the “Multiple Myeloma” disease represents the candidate disease for which the cure and mortality rates will be predicted. The prediction phase is applied on the “COVID-19” virus as well.

4.3 Neural Networks Implementation

The training parameters and configurations of the neural networks have been tuned and adjusted progressively and determined experimentally based on to the most accurate results. Furthermore, in order to ensure a fair comparison, the learning rate for the three models is similar and equal to 0.001 and the number of training steps is equal to 500.

For the DNN and WDNN models, the hidden layers and nodes are defined as following: $h = 1$ hidden layer; $no_1 = 128$ nodes.

4.4 Results

The following subsections include the obtained results for the “Multiple Myeloma” and “COVID-19” candidate diseases, highlighting the predictions quality of the three types of Neural Networks.

As per the below table, and according to the obtained prediction accuracies, the WDNN achieves a better accuracy and outperforms the WNN and DNN, with a significant prediction accuracy equal to 99.9%.

Table 4: Prediction accuracies of the Neural Networks.

“ Multiple Myeloma”							
Actual or true values		WNN Predicted values		DNN Predicted values		WDNN Predicted values	
Cure rate	Mortality rate	Cure rate	Mortality rate	Cure rate	Mortality rate	Cure rate	Mortality rate
81.81%	0%	70.02%	0.06%	79.94%	0.007%	81.79%	0%

	Cure rate Prediction accuracy: 85.6%	Cure rate Prediction accuracy: 97.7%	Cure rate Prediction accuracy: 99.9%
--	---	---	---

Accordingly, the WDNN model is employed to predict the top ten medicines with higher cure rates for the COVID-19 disease, as per the following table. Noting that in the current study, two active ingredients of different medicines are combined at most.

Table 5: Cure rates prediction for several suggested medicines for the COVID-19 disease, using the Wide and Deep Neural Network.

“COVID-19”		
Suggested Medicines	Cure rate %	Mortality rate %
Cyclophosphamide, SAB-301	93.80	0.042045265
Cyclophosphamide, Azithromycin	92.78	0.041968815
Cyclophosphamide, Oseltamivir(Tamiflu)	92.02	0.04189523
Cyclophosphamide, TICAGRELOR	91.57	0.0417286
Cyclophosphamide, Cidofovir	91.27	0.041996222
Cyclophosphamide, Aspirin	91.14	0.041717406
Paracetamol, Cyclophosphamide	90.98	0.041804433
Cyclophosphamide, Phenylephrine	90.89	0.04171965
Cyclophosphamide, GUAIFENESIN	90.84	0.04183637
Cyclophosphamide, Pemetrexed(Alimta)	90.44	0.041622784

Noting that for the suggested one-ingredient treatments, the cure rates of the top ten suggested medicines is predicted as following:

Table 6: Cure rates prediction for several suggested one-ingredient treatments for the COVID-19 disease, using the Wide and Deep Neural Network.

“COVID-19”		
Suggested Medicines	Cure rate %	Mortality rate %
Cyclophosphamide	86.76	0.04068886
SAB-301	80.20	0.03818791
TICAGRELOR	77.98	0.03273421
Cidofovir	77.70	0.03789314
Phenylephrine	77.20	0.03100085
GUAIFENESIN	77.14	0.03313127
Pemetrexed(Alimta)	76.81	0.03075961
Alisporivir	74.48	0.02566667
M7824	74.43	1.3116431
AMODIAQUINE	74.23	0.02041122

5. Conclusion/Discussion

As the COVID-19 pandemic outbreaks all over the world, the development of a new medicine-based treatment or product becomes an urgent need. However, despite the fact that several laboratories are currently conducting clinical studies for possible treatment in existing medicines, estimating and predicting medicines success is a long-term complex task, especially in the context of prediction accuracy.

Nonetheless, there is no powerful and effective medication recommended so far for treating COVID-19, in addition to the possibility of clinical trials failure in the context of unbeneficial and loss-making investments and undesired outcomes. Accordingly, rapidly and urgently exploring the most appropriate medicine-based product for any new pandemic is a fundamental necessity and requires prerequisite steps before proceeding with investments decision in clinical trials and medicine productions.

In this regard, to address this gap, a new quantitative approach has been proposed in the current paper to monitor and evaluate the medicine-based product development process. More precisely, the cure and mortality rates of several potential treatments can be predicted rapidly and systematically using convenient predictive multidimensional neural network techniques and based on the quantitative analysis of Big Data for different indicators or dimensions. More specifically, the main symptoms of several diseases, along with all types of medicines (antiviral medicines, antibacterial drugs, anticancer drugs, etc.) that have been historically attempted in the treatment of different diseases, the active ingredients that compose these medicines, the patents and publications data related to both the disease and the corresponding medicines ingredients, the related reports in clinical trials including the cure and mortality rates of the attempted treatment, etc. can be all considered as major key performance indicators and effective dimensions in evaluating medicine-based product success.

This paper covers two main steps: listing all possible combinations of several medicines' ingredients as suggested single-drug or multiple-drugs treatments, and then forecasting the cure and mortality rates of each proposed or suggested treatment based on the aforementioned dimensions. The present methodology is applied using three different types of Multidimensional Neural Network models for comparison purposes: "Wide model" (WNN), "Deep model" (DNN) and "Wide and Deep model" (WDNN). In addition, this methodology has been experimented on 13 diseases and 28 active ingredients to train the neural networks, then applied on two additional candidate diseases, "Multiple Myeloma" and "COVID-19", for the testing and prediction phases respectively.

The findings show that the Wide and Deep Neural Network model achieves higher prediction accuracy and outperforms WNN and DNN, with a significant prediction accuracy equal to 99.9%, proving to be a valid model that can enhance medicine-based product development.

Therefore, the current findings answer the questions related to the systematical forecast of medicine-based product success using a reliable prediction model. The latter can be employed by decision-makers in clinical medical centers in order to be evaluated at later stages and to validate if the predicted medicines can be considered as a promising and potential treatment. Consequently, addressing this estimation represents a decision support in medicine-based product development for pharmaceutical and clinical medical centers, in order to narrow down and determine their treatment studies areas within the current pandemic.

The proposed approach can be further developed by tackling its current limitations. For instance, the most challenging task faced in the current study is accessing Big Data where publications data are extracted uniquely from Springer and patents data are based only on USPTO granted patents applications. Furthermore, the proposed methodology can be further evolved to include additional diseases and active ingredients and their related clinical trials reports. To this end, the data model proposed in the current paper would need further development in a future work to support further types of medicine-based products.

References

- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, 2014, 7. Retrieved from <http://dx.doi.org/10.1155/2014/614342>
- ALSHEIKH, M. A., LIN, S., NIYATO, D., & TAN, a. H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys and Tutorials*, 16(4), 1996-2018.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning* (1 ed., Vol. 60(1)). New York: Springer.
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhya, H., . . . Shah, H. (2016). *Wide & Deep Learning for Recommender Systems*. Google. Retrieved from <https://arxiv.org/pdf/1606.07792.pdf>
- Ching-Chin, C., Leng, A. I., Ling-Ling, W., & Ling-Chieh, K. (2010). Designing a decision-support system for new product sales forecasting. *Expert Systems with Applications*, 37(2), 1654-1665. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0957417409006034>
- Chow, G. (1995). Portfolio Selection Based on Return, Risk, and Relative Performance. *Financial Analysts Journal*, 51, 54-60. doi:10.2469/faj.v51.n2.1881
- Elaraby, N., Elmogy, M., & Barakat, S. (2016). Deep Learning: Effective Tool for Big Data Analytics. *International Journal of Computer Science Engineering (IJCSE)*, 5, 254-262.
- Fan, A. Y., Gu, S., & Alemi, S. F. (2020). Chinese herbal medicine for COVID-19: Current evidence with systematic review and meta-analysis. *Journal of Integrative Medicine*, 18, 385 - 394. doi:<https://doi.org/10.1016/j.joim.2020.07.008>
- Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary clinical trials communications*, 11, 156-164.
- Foram Panchal, M. P. (2015). Optimizing Number of Hidden Nodes for Artificial Neural Network using Competitive Learning Approach. *International Journal of Computer Science and Mobile Computing*, 358 – 364.
- Gholamreza, S.-F., & Alireza, P. (2018). Design of an artificial intelligence system for predicting success of new product development and selecting proper market-product strategy in the food industry. *International Food and Agribusiness Management Review*, 21(7), 847-864. doi:10.22004/ag.econ.284901

- Golamreza, S., Pooya, A., Kazemi, M., & Naji-Azimi, Z. (2017). Comparison of Multilayer Perceptron and Radial Basis Function Neural Networks in Predicting the Success of New Product Development. *Engineering, Technology and Applied Science Research*, 7, 1425-1428.
- Gosasang, V., Chandraprakaikul, W., & Kiattisin, S. (2011). A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port. *The Asian Journal of Shipping and Logistics*, 27(3), 463-482.
- Levin, G., Dior, U., Shushan, A., Gilad, R., Benschushan, A., & Rottenstreich, A. (2019). Early prediction of the success of methotrexate treatment success by 24-hour pretreatment increment in HCG and day 1–4 change in HCG. *Reproductive BioMedicine Online*, 39, 149 - 154.
doi:<https://doi.org/10.1016/j.rbmo.2019.02.005>
- Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., & Carter, L. J. (2020). Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science*, 6, 315-331. doi:10.1021/acscentsci.0c00272
- Najafabadi, M., Villanustre, F., Khoshgoftaar, T., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep Learning applications and challenges in Big Data analytics. *Journal of Big Data*, 2(1), 1-21.
- National Institutes of Health. (2020, May). *NIH begins clinical trial of hydroxychloroquine and azithromycin to treat COVID-19*. Retrieved from NIH: <https://www.nih.gov/news-events/news-releases/nih-begins-clinical-trial-hydroxychloroquine-azithromycin-treat-covid-19>
- Razmi, M., Hashemi, F., Gheytauchi, E., Manshadi, M. D., Ghods, R., & Madjd, Z. (2020). Immunomodulatory-based therapy as a potential promising treatment strategy against severe COVID-19 patients: A systematic review. *International Immunopharmacology*, 88, 106942.
doi:<https://doi.org/10.1016/j.intimp.2020.106942>
- Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. arXiv preprint arXiv:1609.04747.
- Saade, M., Jneid, M., & Saleh, I. (2019). Predicting technology success based on patent data, using a wide and deep neural network and a recurrent neural network. *International Business Information Management Association (IBIMA 33)*. IBIMA.
- Saade, M., Jneid, M., & Saleh, I. (2020). Enhancing Decision-Making in New Product Development: Forecasting technologies revenues using a Multidimensional Neural Network. *European, Mediterranean and Middle Eastern Conference on Information Systems (17th EMCIS)*. EMCIS.
- Sahadev, V. B. (2020). Predicting numbers of successful new products to launch using soft computing techniques: A case of firms from manufacturing sector industries. *Journal of King Saud University - Computer and Information Sciences*, 32, 254 - 265.
doi:<https://doi.org/10.1016/j.jksuci.2017.08.006>

- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85-117.
- Sun, C.-Y., Sun, Y.-L., & Li, X.-M. (2020). The role of Chinese medicine in COVID-19 pneumonia: A systematic review and meta-analysis. *The American Journal of Emergency Medicine*(0735-6757). doi:<https://doi.org/10.1016/j.ajem.2020.06.069>
- Sun, X., Jiang, J., Wang, Y., & Liu, S. (2020, October). Exploring the potential therapeutic effect of traditional Chinese medicine on coronavirus disease 2019 (COVID-19) through a combination of data mining and network pharmacology analysis. *European Journal of Integrative Medicine*, 40, 101242. doi:<https://doi.org/10.1016/j.eujim.2020.101242>
- USPTO. (2018, june). Retrieved from USPTO: <https://www.uspto.gov>
- Vogenberg, F., Barash, C. I., & Pursel, M. (2010). Personalized medicine: part 1: evolution and development into theranostics. *P & T : a peer-reviewed journal for formulary management*, 35(10), 560-76.
- World Health Organization. (2011). *DEFINITION OF ACTIVE PHARMACEUTICAL INGREDIENT*. World Health Organization. Retrieved from https://www.who.int/medicines/areas/quality_safety/quality_assurance/DefinitionAPI-QAS11-426Rev1-08082011.pdf
- Xiong, H., Dong, Z., Lou, G., Gan, Q., Wang, J., & Huang, Q. (2020). Analysis of the mechanism of Shufeng Jiedu capsule prevention and treatment for COVID-19 by network pharmacology tools. *European Journal of Integrative Medicine*, 40, 101241. doi:<https://doi.org/10.1016/j.eujim.2020.101241>
- Yang, K.-L., Jin, X.-Y., Gao, Y., Xie, J., Liu, M., Zhang, J.-H., & Tian, J.-H. (2020). Bibliometric analysis of researches on traditional Chinese medicine for coronavirus disease 2019 (COVID-19). *Integrative Medicine Research*, 9, 100490. doi:<https://doi.org/10.1016/j.imr.2020.100490>