Journal of Biomedical Informatics 130 (2022) 104075

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Machine learning-based demand forecasting in cancer palliative care home hospitalization

Marzieh Soltani [a], Mohammad Farahmand [b], Ahmad Reza Pourghaderi [c,d,e,*]

[a] Department of Industrial & Systems Engineering, Isfahan University of Technology, Isfahan 841583111, Iran
[b] Department of Electrical & Computers Engineering, Isfahan University of Technology, Isfahan 841583111, Iran
[c] Health Systems Research Center (HSRC), Singapore Health Services, 31 Third Hospital Avenue, Singapore 168753, Singapore
[d] Health Services and Systems Research (HSSR), Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore
[e] Ala Cancer Prevention and Control Center (MACSA), Isfahan 8197113766, Iran

ARTICLE INFO

Keywords:
Management information system (MIS)
Demand forecasting
Home hospitalization
Home care
Cancer palliative care
End of life care
Deep learning
Machine learning

ABSTRACT

Objective: To develop an effective Management Information System (MIS) that is empowered by predictive models that can forecast the demand of end-stage cancer home hospitalized patients in individual and population levels, and help palliative care service systems operate smoothly where the demand is highly fluctuating, resources are limited, expensive, and hardly adjustable in a short time, and the backlog and shortage costs are high.
Method: Inspired by real problems faced by a palliative care center providing various medical, nursing, psychological, and social services in a home-based setting, two Long Short-Term Memory (LSTM) based deep learning models are proposed for demand forecasting at both individual and population levels. The individual-level model can predict the type and time of the next service required for a specific patient with a given demographic and health profile, and the population-level model helps with the prediction of next week's demand for various services in a center supporting a specific patient population. Predicted demand informs on optimal resource and operations plan through a well designed MIS.
Results: Experiments were conducted on a dataset consisting of more than 4000 cancer patients with a Palliative Performance Scale (PPS) of 40 and below discharged from hospital to home under a national palliative care center's home hospitalization service in Iran from September 2012 to July 2019. The models outperformed conventional time-series forecasting methods where applicable. Results indicate that the proposed models were capable of forecasting patients' demand with astonishing performances both individually and on larger scales.
Conclusion: Intelligent demand forecasting can help palliative care home hospitalization systems to overcome the challenge of progressive demand growth when a considerable portion of patients are approaching death, followed by a sudden drop in demand when those patients pass away. It helps to improve resource utilization and quality of care concurrently.

1. Introduction

Tens of millions of people worldwide are affected by life-threatening illnesses such as cancer and HIV/AIDS, which cause them and their families great suffering and economic hardship. The majority of the cases occur in the developing world, where quite often there is little accessibility to prompt and effective treatment for these diseases. The development of palliative care through effective, low-cost approaches is usually the only feasible alternative to respond to the urgent needs of the sick and improve their quality of life [1]. In 2014, the first-ever global resolution on palliative care, World Health Assembly resolution WHA67.19, called upon WHO and the Member States to improve access to palliative care as a core component of health systems, with an emphasis on primary health care and community/home-based care [2]. This paper aims to provide practical tools to support healthcare managers in home-based palliative care settings where the demand is highly fluctuating, resources are limited, expensive, and hardly adjustable in a short time, and the backlog and shortage costs are high.

Several studies have shown that quality of life is greater among patients who die at home compared to patients who die in hospitals, and

* Corresponding author at: Health Services and Systems Research (HSSR), Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore
E-mail address: pourghaderi@u.nus.edu (A.R. Pourghaderi).
URL: https://www.pourghaderi.com (A.R. Pourghaderi).

https://doi.org/10.1016/j.jbi.2022.104075
Received 26 June 2021; Received in revised form 27 March 2022; Accepted 9 April 2022
Available online 28 April 2022
1532-0464/

family members much prefer the patient to die at home [3–7]. Furthermore, many studies have also shown that home-based palliative care is associated with less frequent hospital admissions in the last few months of life, fewer emergency room visits, and a shorter length of stay in hospital [8–15].

This research was initiated to address a real-world challenge faced by a leading cancer palliative care home hospitalization center in Iran, Ala Cancer Prevention and Control Center (MACSA), which provides 24/7 remote counseling and home-based periodic and emergency visits including medical, nursing, psychological, spiritual, social working, and rehabilitation care visits for end-stage cancer patients. Matching supply and demand in this multidisciplinary service center was very challenging in the presence of uncertainty from both patient and service provider sides. Key sources of uncertainty in the subject domain area include urgent requests due to sudden changes in patient's conditions, late cancellations when the patient has been transferred to hospital with serious conditions right before the home care visit that results in increasing the likelihood of death in hospitals, patient's death, and the last-minute patients' unwillingness to receive services. From the service provider side, the center faces unpunctuality and late service cancellation by clinicians and transportation-related uncertainties such as vehicle breakdown, traffic congestion, climate change, and road accident. Other than the aforementioned factors, there is a significant specific source of demand fluctuation in such settings where there is a strong correlation between the demand frequency and death that happens where a typical patient needs more services while approaches death. Fig. 1 illustrates that the number of deaths has a significant statistical correlation with the total volume of medical and nursing care visits in the indexed palliative care center. Therefore the center observes a progressive demand growth when a considerable portion of patients approaching death, followed by a sudden drop in demand when those patients pass away. This makes the ability to predict demand for optimal resource planning a fundamental necessity. Prediction of the future demand from each individual patient enables optimal resource planning and improves the service level to prevent unnecessary hospital admissions and undesired death at hospitals.

There are different quantitative predicting methods that managers can run through their organizations to forecast the demands for services

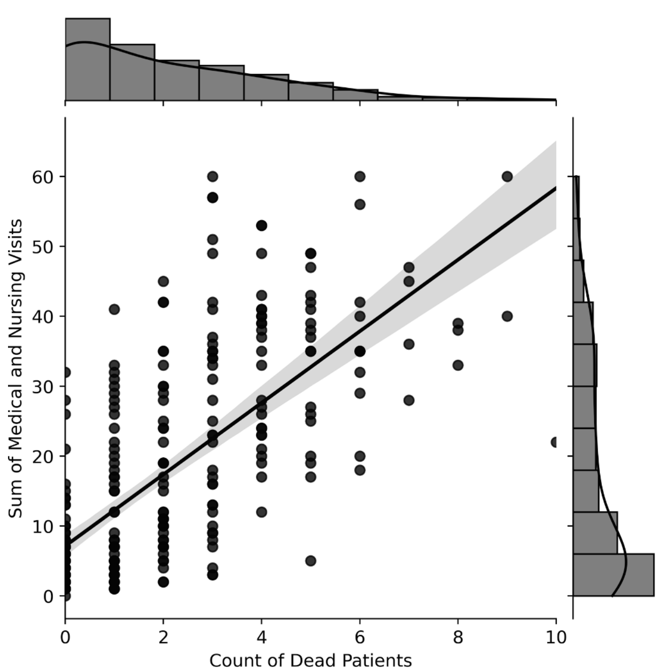

Fig. 1. Joint plot of count of deaths with total volume of medical and nursing care visits, correlation coefficient > 0.7.

in their facilities. Four common simple methods of demand forecasting are percent adjustment, 12-month moving average, trendline, and seasonalized forecast [16]. These four methods are all based upon the organization's recent historical data. Many healthcare organizations often use straightforward techniques for predicting, such as the percent-adjustment method. The other three methods, the 12-month moving average, trendline, and seasonalized forecast, can forecast demand based on historical utilization with higher accuracy. However, due to high dimensional non-linear relations in healthcare recordings, traditional forecasting methods might not be effective for complex time series forecasting problems that involve a large amount of data.

Deep Learning, as one of the most actively researched subjects in recent years, has led to the creation of forecasting models with better performances compared to the traditional ones. Advances in this field have aided data analysts in uncovering complex patterns from sizable amounts of data. Recurrent Neural Networks (RNNs) are one of the branches of Deep Learning algorithms that provide an auto-regression-based mechanism for analyzing sequential data, allowing them to model temporal behaviors far more robustly than traditional neural network architecture. Among this family of algorithms, Long Short-Term Memory (LSTM) networks have been established as state-of-the-art approaches for sequential modeling [17]. Clinical researchers have already applied the LSTM architecture to various problems concerning time series prediction with promising results [18,19] and recently for COVID-19 transmission to support front-line health workers and government policymakers in the pandemic management [20].

In this study, we propose two different LSTM-based neural network architectures, one of which forecasts home hospitalized cancer patients' demands individually and the other on a population level. The first one addresses the key questions of when the next visit for a specific patient with a given demographic and service history profile is going to be requested and what types of care are more likely to be in the patients' needs. This model helps to prevent emergency visit request by each individual patient by scheduling periodic visits and also let the center improve the clinician-patient assignment based on technical considerations and the patient preferences. The second model predicts the total demand size of each service for the next working week of the center based on the current patient population under coverage.

We analyzed the models using the data gathered by Ala Cancer Prevention and Control Center (MACSA). MACSA is a charity organization pioneering non-governmental cancer palliative care in Iran, where the palliative care demand is expected to show an increasing trend due to the aging population [21–23]. It offers a comprehensive package of services in a flexible setting of outpatient clinics, home hospitalization, hospitals, and hospices. Fig. 2 demonstrates the overview of the workflow in MACSA where patients with the Palliative Performance Scale (PPS) of 40 and below benefit from home-based services. PPS is a valid tool to measure physical functional performance in palliative care patients. It assesses five functional parameters, namely degree of ambulation, ability to do activities and extent of disease, ability to do self-care, food and fluid intake, and state of consciousness. PPS are in 10% decrements from 100% (fully ambulatory and healthy) to 0% (death) [24]. It has been found useful for purposes of identifying and tracking potential care needs of palliative care patients, particularly as these needs change with disease progression and also survival prediction [25–27]. PPS is a key input to our predictive models where it was recorded in every visit of an individual patient in MACSA.

The proposed models in individual and population levels will be utilized in the Ala management information system framework to assist their healthcare professionals in improving the quality of care, efficiency, and productivity of this cancer care center by enhancing their management system. The required information such as patients' demographic, medical, and services information in the forecasting process is provided by the information system of Ala, as shown in Fig. 3. By forecasting the required services and their timing for each patient and institute, the planning procedure with two main components of resource

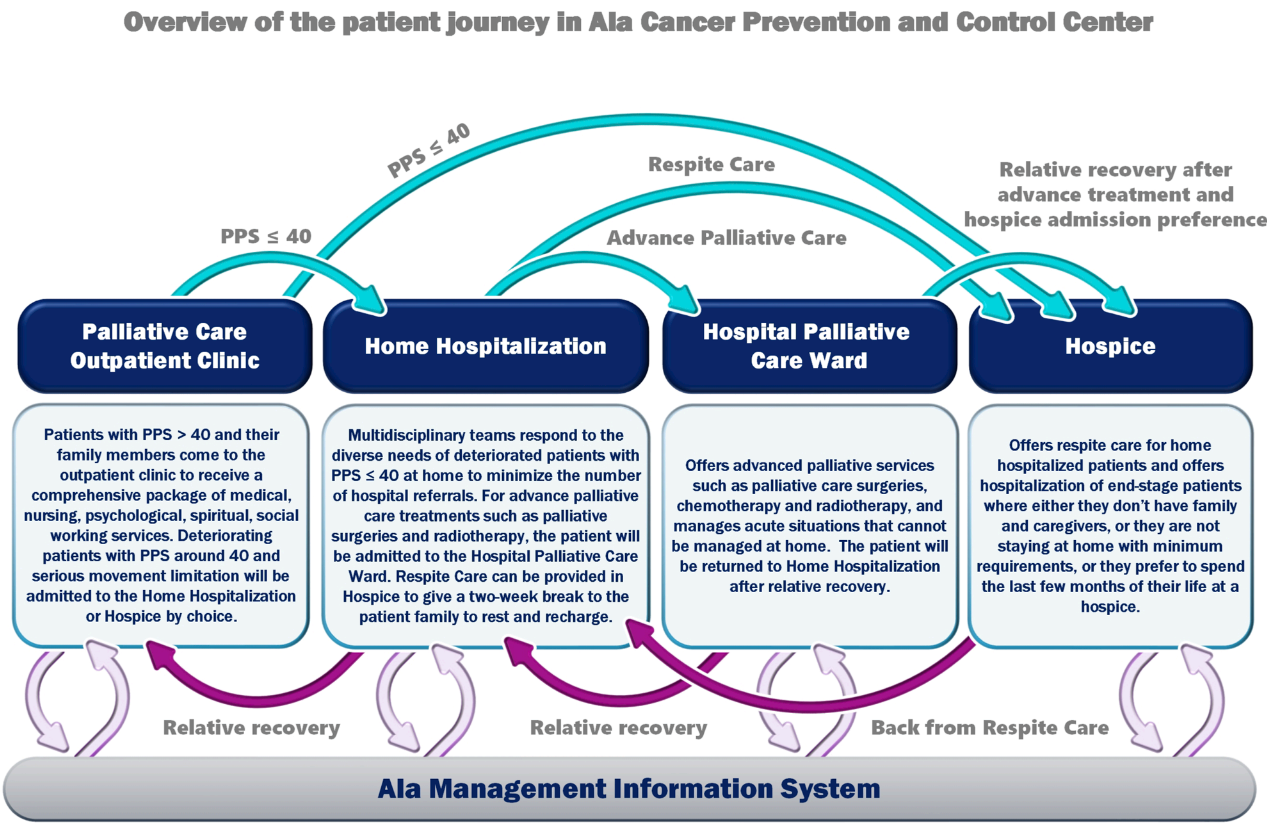

Fig. 2. Overview of the workflow in the indexed palliative care center.

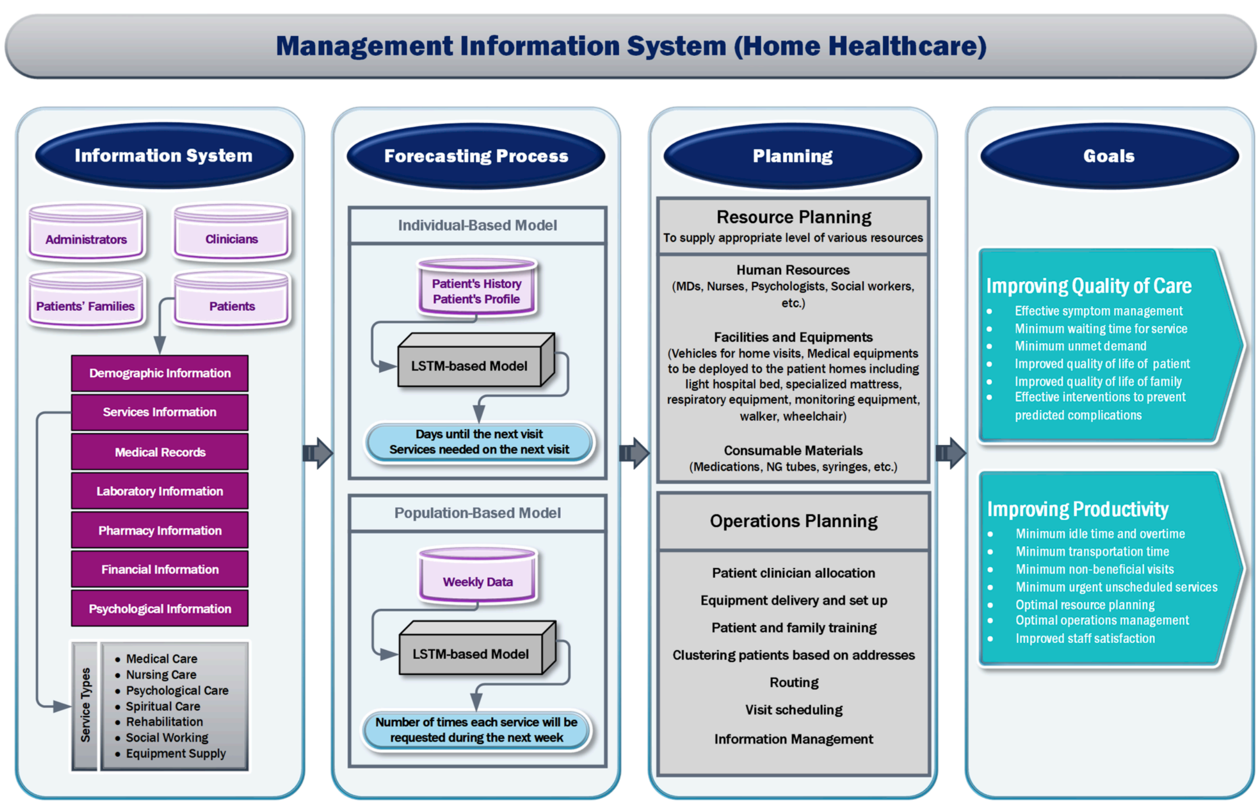

Fig. 3. Management information system of the indexed palliative care center.

planning and operations planning in this system will be optimized to help this cancer prevention and control center operate efficiently.

This research article is organized as follows: Section 2 briefly summarizes previous works related to demand forecasting. 3.1 provides a detailed description and analysis of the most important features of the dataset, 3.2 is a brief introduction for conventional time-series forecasting methods, which we will later use as the baseline for evaluating the proposed models, 3.3 discusses the deep learning models often used when working with sequential data, Section 3.4 provides background for the recurrent neural network design used in our study, and 3.5 follows that by introducing the architecture of the proposed models. In Section 4, the experiments and their results as well as a direct comparison with the baseline auto-regression methods are given. Finally, Sections 5 and 6 discuss the results and concludes the paper, respectively.

2. Related Works

A time series is a sequence of data points related to a variable that changes over time. Health care datasets are sequences of observations taken successively in time which is one of the thriving sources for patients' demand forecasting. Understanding the pattern of patients' demands can aid health care organizations in resource management and hence improves the overall efficiency of the system. Accordingly, Traditional time series prediction techniques, such as the Auto-Regressive (AR) technique and its variants, have been extensively used for demand forecasting. Several researchers have studied the necessity of forecasting daily patient volumes in hospital emergency departments (ED). Jones et al. focused on forecasting daily patient volumes in the emergency department using ARIMA, regression, and neural network techniques [28]. In another research, Jones et al. studied the temporal relationships between the demands for key resources in the ED and claimed that results suggest that multivariate time series models can be used to forecast ED patient census reliably; however, forecasts of the demands for diagnostic resources were not sufficiently reliable to be useful in the clinical setting [29]. Traditional time series sequencing techniques also have been applied to predict ED crowding [30–32].

Conventional Time Series Forecasting Methods focus on univariate data with linear relationships and fixed and manually diagnosed temporal dependence; such methods have trouble forecasting when the time series start taking characteristics such as seasonality, trends, and auto-correlation simultaneously. As an instance for time series data modeling with periodic characteristics, if the patient goes through different stages of the illness, the AR model cannot change and adapt to the new demands of the patient due to its static processing nature. As the statistical properties of a time series such as mean, variance, etc., remain constant over time, the time series would be considered stationary.

ANNs are one of the most robust and widely used prediction methods in many areas as they have been found very efficient in finding patterns in non-stationary and nonlinear time series. As practical and flexible modeling tools, ANNs can generalize pattern information to new data, tolerate noisy inputs, and produce reliable and reasonable estimates [33]. ANNs are successfully applied to develop a risk advisor model to predict the chances of diabetes complications according to the changes in risk factors [34]. They have also been used to identify the optimal subset of attributes from a given set of attributes for the diagnosis of heart diseases and to predict heart disease with a good performance [35]. ANNs were also applied in modeling daily patient arrivals in the Emergency Department [36] and diagnosis of disease based on age, sex, body mass index, average blood pressure, and blood serum measurements [37]. ANNs have shown to outperform statistical models in the diagnosis of coronary artery disease with 84.7% accuracy and 86.5% positive predictive rates in the detection of the illness by utilizing RNN layers [38]. Purwanto et al. used three types of data in healthcare, namely, infant mortality rate, the morbidity of tuberculosis, and crude birth rate, and developed hybrid linear regression-neural network forecasting models for health departments since good and accurate forecastings are very helpful in devising appropriate action plans [39]. To the best of our knowledge, this study is the first application of ANN in home hospitalization demand forecasting.

3. Material and Methods

3.1. Data and Features

The dataset contained details of 12358 cancer patients, 3837 of whom are home hospitalized patients served by MACSA from September 2012 to July 2019. Each patient profile included information of patient ID, registration date, date of birth, gender, marital status, province, city, alive-dead status, death date, nationality, cancer type, pathology grade, pathology metastasis, and cancer staging and also time series data on the service history that includes service date, service type and palliative performance scale (PPS) associated with each service. Data were anonymized and cleaned in order to collect and analyze applicable data about home hospitalized cancer patients. The number of patients reached 743 after cleaning, where we only considered expired patients with complete journeys in the home hospitalization department.

Fig. 4 shows the most prevalent types of cancer among home care cancer patients categorized by gender and alive-dead status. As can be seen, the code of C50.9, which refers to the malig-t neoplasm of breast of unspecified site, is the most prevalent cancer among all cancers where the malig-t neoplasm of the brain of unspecified sites with the ICD-10-CM code of C71.9 is the most common cancer in males.

Fig. 5 presents the total demand size for each service type by home hospitalized patients during the study period. As is shown in Fig. 5, the demand size for medical and nursing care visits are the highest, followed by rehabilitation service and psychological care services, and then the spiritual care, social works, and finally supplying home care equipment such as hospital bed package with electric mattress, oxygen machine, suction unit, and wheelchair.

Patients' ages range from 3 to 95, and the male gender forms about 52% of the dataset. As is shown in Fig. 6, most of the patients' population is between 60 and 90 years old. This Figure also illustrates that the most frequent ages among patients ranging from 54 to 78 are almost equal in terms of gender.

The stage of cancer defines the size of a tumor and how advanced the

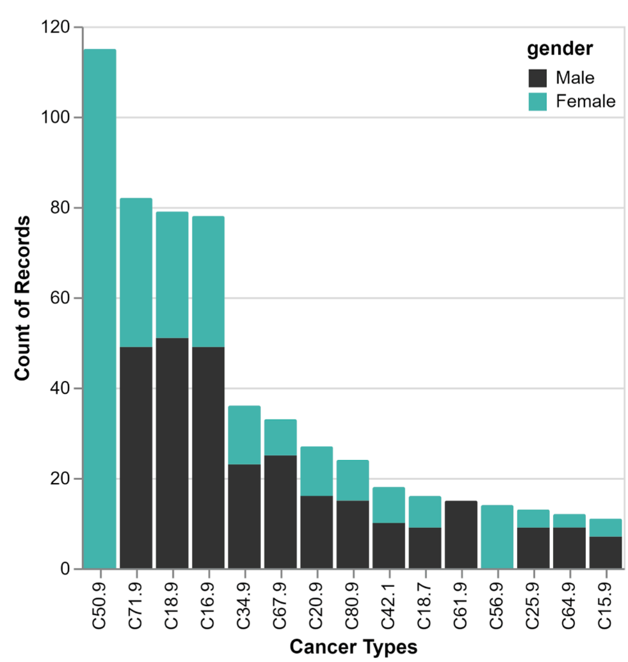

Fig. 4. Counts of home hospitalized patients for the most prevalent ICD-10-CM codes.

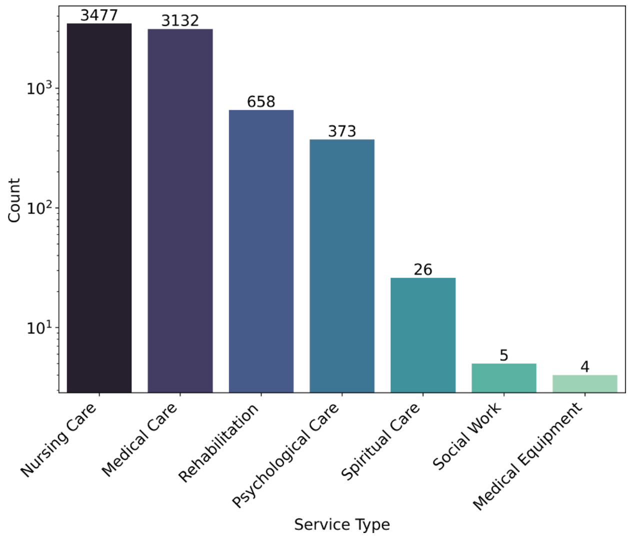

Fig. 5. Service counts for home hospitalization services 2012–2019.

cancer is. Fig. 7 demonstrates that patients with the most frequent cancer types are mainly in stage 4, which means the cancer is more extensive and may have spread to the surrounding tissues or the lymph nodes, also called secondary or metastatic cancer. Patients with C42.1, the Hematopoietic, and the Reticuloendothelial System of Bone Marrow are mainly in stage 5, which means cancer has spread to at least one other body organ. It is also known as advanced or metastatic cancer.

3.2. Auto-Regression Methods for Time-Series Forecasting

Auto-Regression (AR) models are the conventional techniques for time-series forecasting [40]. This family of methods incorporates various models that progressively increase in their modeling *capacity*; their ability in capturing certain attributes of a time-series. The most important attributes that a time-series can have are level, trend, and seasonality. Furthermore, while some models require the data to be weakly stationary, some others are able to integrate a difference mechanism into their formulation to internally handle this undesired property. On the other hand, most AR methods are single-variate, while some others can take in and forecast multiple variables at the same time, which might improve the results. Last but not least, this family of algorithms suffer from a major shortcoming, which is that the samples in the time-series an AR method can model must be equally spaced in time. This is not a luxury most problems enjoy.

In our study, we attempt to compare our proposed models against a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model, arguably the most powerful single-variate AR method capable of capturing all the aforementioned attributes and properties of a time-series, and Vector Auto-Regression (VAR), a well-established multivariate AR model that can simultaneously process and forecast multiple related variables. Both these models contain several hyperparameters, *nobs* on the model that need to be tuned for optimal results. These hyperparameters are associated with the attributes and properties the model must capture, and tuning them is often accomplished through trial and error or empirical analysis of the time-series and looking for the said attributes and properties.

3.3. Deep Learning and Sequential Data

At the heart of Deep Learning lie the artificial neural networks, statistical models capable of capturing patterns within data. A neural network is expected to model an arbitrary function, $f : X \rightarrow Y$, taking X as its input and outputting values as close to Y as possible. In a supervised setting, neural networks achieve this by *training* on a dataset of X-Y pairs, a process during which *learnable parameters*, placed throughout their structures, are fine-tuned to minimize some measure of the error in the outputs. The mathematical function used to measure the error is referred to as the loss function of the model. If we denote this function as J, the function modeled by the neural network as $\widehat{f}$, the outputs of the network as $\widehat{Y}$, and the set of learnable parameters in the network as Θ, the training process is essentially solving the optimization problem described in Eq. 1.

$$\min - \Theta \quad J(Y,\widehat{Y}) = J(Y,\widehat{f}(X)) \tag{1}$$

The optimization algorithm searches for the optimal parameter set in an iterative manner, readjusting the weights to reduce the error with every step. However, the training process cannot converge to a set of optimal parameters if the network structure does not provide it with enough capacity to capture the patterns in the data. On the other hand, if the

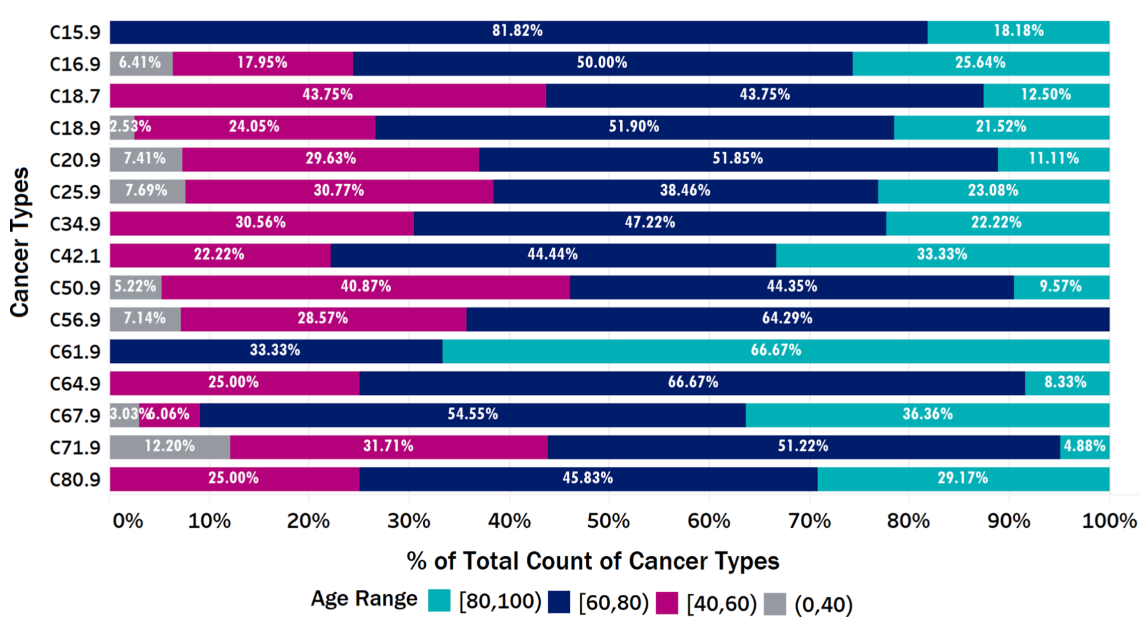

Fig. 6. Stacked bar chart of the most prevalent cancer types stratified by age.

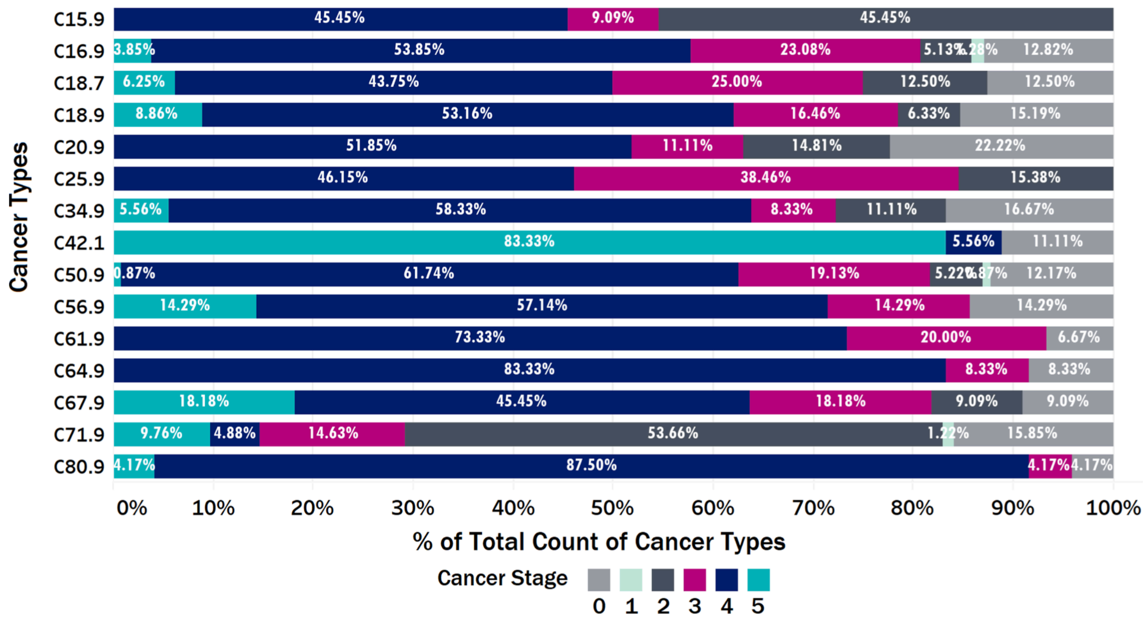

Fig. 7. Distribution of the stages in the most prevalent cancer types.

neural network structure is more complicated than necessary, the resulting model is very likely not to generalize well and perform poorly on new data. Thus, crafting the proper structure for the network is basically a trial and error guided by some intuition. The most basic neural networks consist of several layers, possibly of different types, which are often stacked on top of one another, processing the input values and then passing the results to the next one in line. The layers contain varying numbers of neurons, the inner structures of which depend on the layer's type.

Feedforward layers, which model a vector-to-vector transformation, are the first and most basic type of layers devised. Let us denote the n-D input vector of a feedforward layer with u neurons as x. The output of this layer, which will be denoted as h, is a u-D vector, the elements of which are the outputs of its individual neurons. Eq. 2 shows the exact formulation for the output of the neurons.

$$h - j = \sigma\left(w^j \cdot x + b^j\right) \tag{2}$$

In Eq. 2, w^j, which is an n-D vector, and b^j, which is a scalar, are the learnable parameters of the neuron that will be later fine-tuned. The function denoted as σ in this equation is known as the *activation function* of the layer. Given the fact that the input of this function is a linear combination of the elements of x, a non-linear σ has to be used if the feedforward layer is to model a non-linear transformation. Typical choices for σ are the hyperbolic tangent, logistic sigmoid, and the Rectified Linear Unit (ReLU) functions [41].

While powerful, feedforward layers are simply ill-suited when we are dealing with sequential data. Recurrent Neural Networks (RNNs) have been at the forefront of deep learning when it comes to sequential data. Recurrent layers are essentially feedforward layers with *memory*. Variants of recurrent layers are able to access their own outputs at previous time-steps through a variety of different mechanisms. The simplest recurrent layer, a *vanilla recurrent layer,* is applied to the elements in the input sequence one after another. During this process, each recurrent layer remembers its output at the previous time-step and takes it in as another input.

Eq. 3 formulates this behavior of recurrent neurons.

$$h - j^t = \sigma\left(w - h^j \cdot h - j^{t-1} + w - x^j \cdot x^t + b^j\right) \tag{3}$$

In Eq. 3, $w - h^j, w - x^j$, and b^j represent j-th neuron's learnable parameters. On the other hand, x^t denotes the input vector at time-step t, while h^t and h^{t-1} denote the outputs of the layer at the same time-step and its predecessor. Note that the learnable parameters are shared across the time-steps. Vanilla recurrent layers fail to capture temporal patterns that span over long durations. Furthermore, the formulation of these neurons is prone to crippling gradient problems, where the gradient signals used to fine-tune the parameters either go to zero or reach unusually high values, halting or destabilizing the training process. Various alternatives to vanilla recurrent layers have been proposed to address these issues.

While popular, RRNs are not the only widely adopted solutions for sequential data in deep learning. Influenced by their immense success in computer vision, researchers have tried to apply convolutional neural networks (CNNs) to various problems. This variant of CNNs, which has been dubbed 1-D CNN, has been remarkably successful [42]. However, there are certain drawbacks in their design that has prevented them from becoming the primary tool for modeling sequential data. The most important issue in 1-D CNNs is perhaps their lack of a *memory* mechanism, which otherwise could have enabled them to remember useful features over long distances in sequences. Furthermore, most implementations of 1-D CNNs pose a lower-bound on the length of input sequences. This factor alone prevents us from applying this model to the problems discussed in this study.

Another major innovation in this domain is the attention mechanism and its utilization in transformers [43]. These models have revolutionized natural language processing, one of the primary fields that deals with sequential data. However, it is important to note that successful transformers are often resource extensive, and researchers are actively trying to reduce their processing demands [44]. Even with this issue, transformers remain an option for the problems discussed in our study. However, we opted to use a powerful and robust variant of RNNs that is significantly less demanding when it comes to processing resources, and yet, it manages to achieve excellent results in numerous studies.

3.4. Long Short-Term Memory

As previously mentioned, vanilla recurrent layers suffer from problems that hinder the learning process. Researchers have proposed new variants that attempt to rectify these issues. Long Short-Term Memory (LSTM) [17] is the most complex and perhaps the most successful RNN variant that has been proposed [45]. Naturally, it is the best choice for the task at hand. An LSTM neuron is significantly more complicated than a vanilla recurrent neuron, containing an extra memory unit and several *gates*over the flow of information. Each information gate is similar to a vanilla recurrent unit which mixes the input and the output at the previous time-step for a specific purpose. Let us consider an LSTM layer with k neurons. In this case, if we denote the input vector at time-step t as x^t, the output of the layer is calculated as in Eq. 4. Fig. 8 illustrates the same calculations.

$$\begin{aligned} f^t &= \sigma - g\left(W - fx^t + U - fh^{t-1} + b^f\right) \\ i^t &= \sigma - g\left(W - ix^t + U - ih^{t-1} + b^i\right) \\ o^t &= \sigma - g\left(W - ox^t + U - oh^{t-1} + b^o\right) \\ \widetilde{c}^t &= \sigma - c\left(W - cx^t + U - ch^{t-1} + b^c\right) c^t = f^t \odot c^{t-1} + i^t \odot \widetilde{c}^t h^t = o^t \odot \sigma - h(c^t) \end{aligned} \tag{4}$$

In Eq. 4, c denotes the new memory unit of the neurons, which is often referred to as their hidden state. $\sigma - g, \sigma - c$, and $\sigma - h$ are the activation functions of the gates, the hidden variables, and the output vectors, respectively. In [17], the authors suggest the logistic sigmoid for $\sigma - g$ and the hyperbolic tangent for $\sigma - c$ and $\sigma - h$. f^t is known as the forget vector of the neuron (at time-step t), which regulates the effect of the previous hidden state value on the next. The newly generated hidden state is then regulated using i^t, the input vector at time-step t. The two regulated flows are then added to form the hidden state at time-step t, a design choice closely linked to the gradient problems in the previous iterations of recurrent neurons. The hidden state goes through another gate, known as the output gate of the neuron. The output vector generated by this gate, which is denoted as o^t, is applied to an activated hidden state and then passed to the next time step.

More often than not, LSTM layers and feedforward layers are utilized together in a time series prediction problem. A typical setting is to first pass the sequence of vectors through a few LSTM layers to recognize and extract the temporal patterns within the data. Once all the vectors in the sequence have been processed, the output vector at the last time-step is taken and passed through several feedforward layers, which can then map the extracted temporal features to the desired predictions. We will be using a similar configuration in our models.

3.5. Model Design

3.5.1. Individual-Based Model

Fig. 9 illustrates the design and structure of the individual-based model. The neural network takes in the patient's profile, details about them and their disease, as well as a sequence of vectors recounting all their previous visits. The features included in the patients' profiles and histories can be divided into to main categories: the discrete and qualitative features, including their gender, marital status, province of residence, disease (expressed in the form of ICD-10-CM codes), cancer stage, and grade, and provided services, and the quantitative features, age, metastatic, days in-between visits, and patients' PPS at the time of visit. The first category of features is numericized in the form of one-hot vectors, while the features in the second category are merely rescaled to fit on the $[0,1]$ range. Values that do not lie in a small neighborhood around zero can destabilize a neural network's training process. Therefore, a common practice in Deep Learning is to normalize the features.

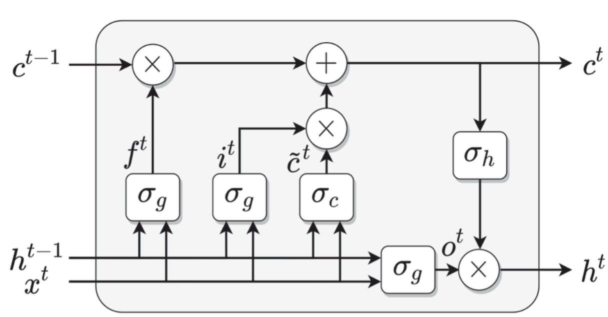

Fig. 8. The structure of an LSTM neuron. An LSTM layer has to memorize its output (denoted as h^{t-1}) as well as its hidden state (denoted as c^{t-1}) at the previous time-step.

In the lowest levels of the network, a pair of LSTM layers and a pair of feedforward layers are used to extract features from the patients' histories and profiles. The extracted features, which come in the form of two vectors, are then concatenated to one another before being fed into two separate pairs of feedforward layers, which are meant to extract the features to the desired predictions: days in-between the patients' latest visit and their next one, and the services they will need on their next visit. The LSTM layers use the activation functions suggested by [17]. As for the feedforward layers, the Rectified Linear Unit (ReLU) activation function has been used on all the layers, except for the layers that are meant to output the prediction values (No. 4 and 6 in Fig. 9), which use a linear function and the logistic sigmoid function for outputting the days and the services, respectively. To ensure the generalizability of the models, a mechanism known as Dropout [46] has been applied to the output of the first layer in each feedforward pair. This mechanism comes with a *hyperparameter*, known as its rate and denoted as α, which needs to be carefully tuned through trial and error, similar to the network's overall structure.

An important factor to consider, which complicated the learning process for the network, is that the sequences of vectors, corresponding to patients' previous visits, come in various lengths, ranging from 2 to around 150 data points. To resolve these complications, we employed a technique known as curriculum learning [47], which can significantly improve learning in recurrent neural networks [48]. In our implementation of this technique, we slowly increase the length of the sequences the network observes as the training progresses. This growth is controlled by a hyperparameter which we denote as α. If i represents the iteration of training, we are currently at the maximum length for the sequences the network trains on is calculated using Eq. 5.

$$\text{Length} = \alpha \cdot \text{ceil}\left(\sqrt{i}\right) \tag{5}$$

This formulation gradually increases the length of the sequences the network has to process but also increases the number of iterations between each increase in length, giving the network more time to learn the patterns in sequences of longer lengths.

3.5.2. Population-Based Model

The population-based model shifts the focus of the model to the demands on a higher level. The slightly simpler structure of this new model is illustrated in Fig. 10. The inputs and the outputs of the network are all discrete values, which we rescaled using their mean and standard deviation so that the distribution for each of the services approximates a normal distribution around zero. The new structure consists of two LSTM layers, followed by three feedforward layers with activation functions similar to those of the individual-based model. We also employed the dropout mechanism, but only on the output of the first feedforward layer. Considering the applications this model can have, it makes sense to use fixed-length sequences as the inputs for this model. As a result, there is no longer a need for utilizing curriculum learning in training.

4. Experiments and Results

Evaluation of the models requires unseen data to ensure that the learned features generalize to cases the network has not encountered during training. A typical practice is to split the dataset into three subsets, one used for training the model, one for tuning its hyperparameters, and one for the final evaluation of the tuned model. However, in cases where one is dealing with a small dataset, the latter two are often combined. We followed this approach and used 80% of the data for training the model, leaving 20% aside as our validation set. We trained a single individual-based model for all the patients in a palliative care

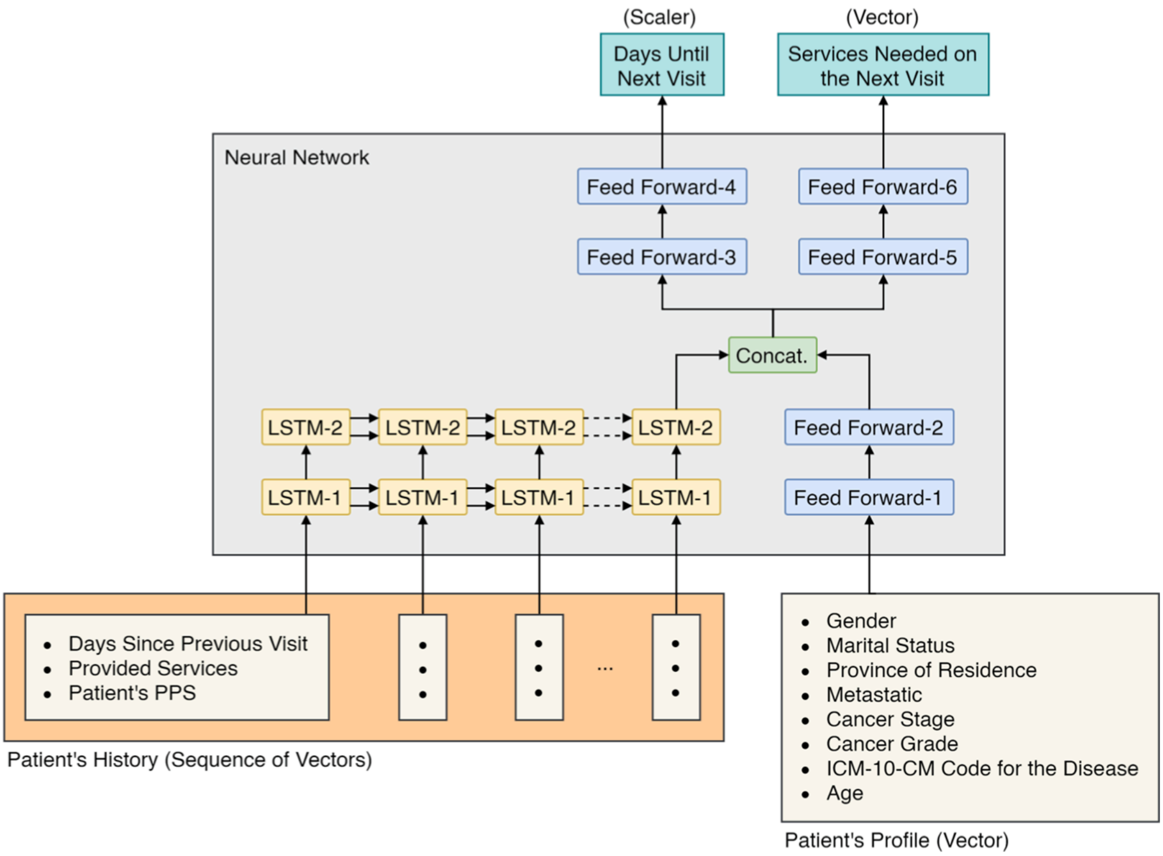

Fig. 9. Design of the individual-based model. The temporal features extracted by the pair of LSTM layers are concatenated to the features extracted from the patient's profile, before being mapped to a scaler and a vector through two separate sets of feedforward layers.

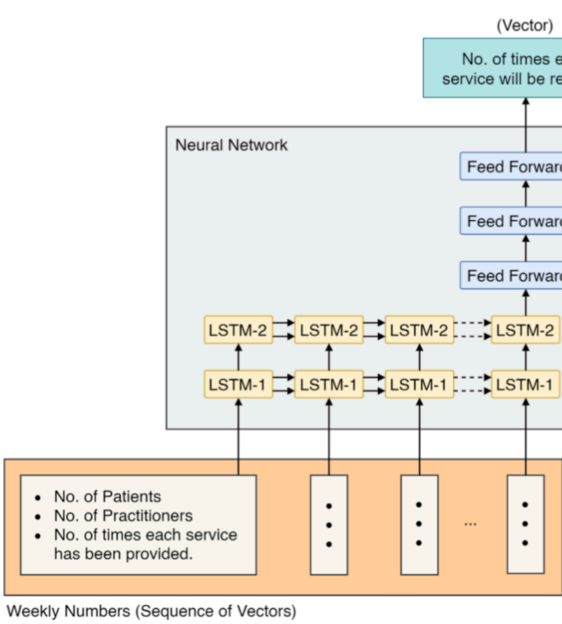

Fig. 10. Design of the population-based model. The structure of the model is slightly simpler, compared to the model used for individual-based predictions. However, experiments show that this network requires far more neurons with learnable parameters to accomplish its task.

home hospitalization program. However, the population-level sequences extracted from the dataset for the two institutes differed greatly in their trends and seasonalities. As a result, the two institutes were modeled separately using two identical population-based networks.

Due to the limitation discussed in 3.2, we can only compare the population-based model to an AR-based technique, as a baseline, since the temporal distance between datapoints in individual-based sequences vary. We use the implementations provided by the statsmodels library for Python language. We first model the sequences associated with individual services in each institute as single-variate time-series using independent SARIMA models. We then model all the services at once using a single VAR model for each institute. To tune the hyperparameters on the SARIMA models, we used the Bayesian optimization strategy, quickly searching through over 7000 combinations of hyperparameter values. In the case of VAR models, the lag order of the model, its single hyperparameter, was tuned automatically using the Akaike information criterion.

Table 1
Hyperparameter values chosen for the model.

Layer Name	Neurons	Dropout Rate
	Individual-Based Model	
LSTM-1	25	-
LSTM-2	25	-
Feedforward-1	25	0.5
Feedforward-2	25	-
Feedforward-3	25	0.5
Feedforward-4	1 (Outputs scaler)	-
Feedforward-5	25	0.5
Feedforward-6	8 (No. of Services)	-
	Population-Based Model	
LSTM-1	128	-
LSTM-2	128	-
Feedforward-1	64	0.50
Feedforward-2	32	-
Feedforward-3	8 (No. of Services)	-

Table 1 enlists the hyperparameters related to the structure of the network. We used the Mean Squared Error as the loss function for the quantitative outputs and the Binary Cross-Entropy for the prediction of services in each visit. We implemented the models using the PyTorch library and run the experiments on an NVIDIA 1660 Ti graphic card. The models are all trained using the Adam [49] optimizer, with a learning rate of 10^{-3}, and its $\beta-1$ and $\beta-2$, two crucial hyperparameters of this algorithm, to 0.9 and 0.999, respectively.

In order to evaluate the individual-based model on how well it can predict future services, typical metrics for multi-label classification problems can be applied. However, an essential factor to consider is that the services are not balanced regarding how often they are demanded; some services are provided on nearly every visit, while others have only been registered a handful of times during the course of treatment. With this imbalance present in the data, typical metrics such as the model's prediction accuracy, formulated in Eq. 6, can not be a good indicator of its performance. As a result, we adopted the F_1 score evaluation metric, the harmonic mean of precision and recall, which can provide a better depiction of the model's performance in case of an imbalanced dataset. Precision and recall themselves are calculated using Eqs. 7 and 8, where TP, TN, FP, and FN stand for True Positives, True Negatives, False Positives, and False Negatives respectively. Precision provides us with a measure of how confident we can be that a positive prediction by the model is actually a positive case. On the other hand, recall calculates how likely the model is to capture all the positive cases.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

The individual-based model's outputs for the services are probability values lying in the $[0, 1]$ range, indicating how likely the patient will require that service. These probabilities are transformed into definitive states with respect to a 0.5 threshold. Table 2 contains the values for the classification metrics, calculated separately for each service.

For the performance evaluation of the models on their quantitative predictions, namely, the days in-between visits and the number of weekly services, we calculate the Mean Absolute Error (MAE) of the predictions, a typical metric for regression models, as well as a confidence interval for the mean difference between the predicted values and the actual numbers. The results can be found in Table 3.

5. Discussion

Demand forecasting continues to be a top priority in healthcare settings, where managers have aligned incentives to improve decision-making processes to ameliorate the quality of care patients receive and formulate adequate strategies to overcome most challenges such as limited resources, shortage of workforce, and budget allocation. Forecasting patients' demands in healthcare facilities can help administrators and professionals improve planning by utilizing predictive models in their management information systems to provide the best possible services by making the best use of the available resources. The proposed LSTM-based models in this research can effectively improve the optimization procedures in resource and operations planning to achieve the best possible quality of care. This management information system framework allows all involved decision-makers to identify patients' needs and effectively plan to address their demands.

Demand forecasting enables home healthcare managers to optimally addresses many key challenges in home-based healthcare resource planning and operations management. Manpower planning is complicated where we can consider different forms of contracts including per visit payment with a fixed payment for every booked time block (a half-day shift in the indexed center), weekly agreements, monthly agreements, annual agreements, and permanent agreements from the lowest to the highest cost-effective forms. The proposed management information system can support managers to find the optimal number of staff from various disciplines to be hired with different forms of contracts. The decision to open new positions was made based on the projected demand every year and there is a list of potential collaborators that be asked to join for a day, week, or month to cover the demand surges. Optimal human resource planning can balance the trade-off of staff idle time and patients waiting time for services and unmet demands.

Inventory management for both durable and consumable equipment is another challenging task that can be performed more productively using the proposed MIS. Durable medical equipment like hospital beds is stored in central storage and transferred to the centre's equipment supply department storage based on the predicted demand in a weekly manner. Procurement planning for medications and consumable items like NG tubes could be guided by demand forecasting modules. Another challenging decision is related to fleet sizing where managers should hire vehicles on an hourly, daily, weekly, monthly or annual basis. Effective prediction of the number of visits can inform on the optimal transportation plan.

Prediction of the next required service for each patient allows optimal clinician-patient allocation where there are preferences from both patient and service provider sides. A very important service that most patients and families greatly appreciate is the availability of care through in time services in the last few days before and at the moment of death at patient homes. When a significant proportion of the patients approach to death, the demand will progressively increase and providing in time services would be extremely challenging. It becomes more intractable when the center faces sudden drops in demand when those deteriorating patients pass away.The proposed MIS enabled the service center to efficiently respond to demand peaks.

We described and evaluated two LSTM models to forecast the demand for a range of palliative care services who are hospitalized at home with advanced cancer. The individual-based model forecast for a patient when that patient will need a service and what types of services will be needed at that time. The population-based model forecasts for a service center the quantity of services that will be requested in the coming week. The proposed deep learning models were shown to be capable of

Table 2
Performance metrics values for the predictions made by the individual-based model for the services demanded on the patient's next visit. Total Samples: 607.

Service ID	Service Type	Number of positive services	Accuracy	Precision	Recall	F_1 Score
2	Medical Care	392	63.76%	64.78%	96.17%	77.41%
3	Nursing Care	438	68.70%	73.48%	88.58%	80.33%
8	Psychological Care	58	90.61%	100.00%	1.72%	3.39%
10	Rehabilitation	89	85.34%	0.00%	0.00%	0.00%
18	Spiritual Care	5	99.18%	0.00%	0.00%	0.00%
9	Social Work	0	100.00%	0.00%	0.00%	0.00%
7	Psychological Biography	0	100.00%	0.00%	0.00%	0.00%
12	Medical Equipment	2	99.67%	0.00%	0.00%	0.00%
	Weighted Average		69.75%	64.41%	77.84%	66.80%

Table 3
Evaluation of the quantitative predictions of the models.

Center	Variable		Model	MAE	Prediction Errors Confidence Interval 90%	95%	99%
1 & 2	Days in-betweenVisits		Individual-Based	13.4952	-2.59431, 0.53323	-2.89476, 0.83368	-3.48337, 1.42229
1	2	Medical Care	Population-Based	2.6406	−0.55723, 1.12763	−0.72401, 1.29441	−1.05881, 1.62921
			SARIMA	5.2105	−0.61465, 2.35086	−0.90784, 2.64406	−1.49586, 3.23207
			VAR	10.980	9.11458, 12.08314	8.82108, 12.37664	8.23246, 12.96526
	3	Nursing Care	Population-Based	3.0655	−1.05020, 0.81832	−1.23516, 1.00328	−1.60645, 1.37457
			SARIMA	5.5923	−2.84333, 0.39698	−3.16369, 0.71735	−3.80620, 1.35985
			VAR	14.267	12.31527, 15.66215	11.98437, 15.99305	11.32073, 16.65669
	8	Psychological Care	Population-Based	0.8497	−0.21413, 0.28816	−0.26385, 0.33787	−0.36366, 0.43768
			SARIMA	1.5440	0.21325, 1.07732	0.12782, 1.16275	−0.04351, 1.33408
			VAR	2.2848	1.68089, 2.53859	1.59609, 2.62339	1.42603, 2.79346
	10	Rehabilitation	Population-Based	1.7918	−0.52376, 0.54802	−0.62985, 0.65412	−0.84283, 0.86709
			SARIMA	2.0879	−0.04924, 1.39861	−0.19238, 1.54175	−0.47947, 1.82884
			VAR	2.1749	−0.38370, 1.06725	−0.52716, 1.21070	−0.81486, 1.49840
	18	Spiritual Care	Population-Based	0.2084	0.00838, 0.17552	−0.00817, 0.19206	−0.04138, 0.22527
			SARIMA	0.1053	0.03667, 0.17385	0.02311, 0.18742	−0.00409, 0.21462
			VAR	0.1814	−0.05933, 0.07823	−0.07293, 0.09183	−0.10021, 0.11910
	9	Social Work	Population-Based	0.0177	−0.02054, −0.01230	−0.02135, −0.01149	−0.02299, −0.00985
			SARIMA	0.0351	−0.02360, 0.09377	−0.03520, 0.10538	−0.05847, 0.12865
			VAR	0.0429	−0.03170, 0.08564	−0.04331, 0.09724	−0.06657, 0.12050
	7	Psychological Biography	Population-Based	0.0063	−0.00466, −0.00039	−0.00508, 0.00003	−0.00593, 0.00088
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0036	−0.00375, −0.00317	−0.00381, −0.00312	−0.00392, −0.00300
	12	Medical Equipment	Population-Based	0.0287	−0.02133, 0.03921	−0.02733, 0.04521	−0.03936, 0.05724
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0111	−0.01172, −0.01054	−0.01184, −0.01043	−0.01207, −0.01019
2	2	Medical Care	Population-Based	2.4161	−1.16637, 0.94326	−1.38472, 1.16161	−1.83935, 1.61624
			SARIMA	2.5619	−2.09086, 0.02088	−2.30808, 0.23810	−2.75802, 0.68804
			VAR	2.6098	−0.18268, 1.95927	−0.40300, 2.17959	−0.85938, 2.63597
	3	Nursing Care	Population-Based	1.7358	−1.03603, 0.62834	−1.20830, 0.80061	−1.56697, 1.15928
			SARIMA	2.0559	−0.48004, 1.25296	−0.65830, 1.43122	−1.02754, 1.80047
			VAR	2.4148	0.62804, 2.32400	0.45359, 2.49844	0.09224, 2.85980
	8	Psychological Care	Population-Based	0.6018	−0.17154, 0.33836	−0.22432, 0.39114	−0.33421, 0.50102
			SARIMA	0.4465	0.03188, 0.47846	−0.01406, 0.52439	−0.10921, 0.61954
			VAR	0.5606	−0.11211, 0.32345	−0.15691, 0.36826	−0.24971, 0.46106
	10	Rehabilitation	Population-Based	0.0623	−0.05292, 0.00278	−0.05869, 0.00854	−0.07069, 0.02055
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0063	−0.00611, −0.00262	−0.00646, −0.00226	−0.00721, −0.00152
	18	Spiritual Care	Population-Based	0.0038	−0.00119, 0.00191	−0.00151, 0.00223	−0.00218, 0.00290
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0000	–, –	–, –	–, –
	9	Social Work	Population-Based	0.0011	0.00026, 0.00099	0.00019, 0.00106	0.00003, 0.00122
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0000	–, –	–, –	–, –
	7	Psychological Biography	Population-Based	0.0001	−0.00001, 0.00008	−0.00002, 0.00009	−0.00004, 0.00011
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0000	–, –	–, –	–, –
	12	Medical Equipment	Population-Based	0.0004	−0.00012, 0.00021	−0.00016, 0.00024	−0.00023, 0.00032
			SARIMA	0.0000	–, –	–, –	–, –
			VAR	0.0063	−0.00611, −0.00262	−0.00646, −0.00226	−0.00721, −0.00152

demand forecasting in both individual and population levels. The individual-based model has an average accuracy of 69.75% and F1 score of 66.8% in predicting the types of services needed by a patient. The population-based model has average MAE of predictions for institutes No.1 and No.2 are 1.07 and 0.6, respectively. Furthermore, we compared the performance of the model against two conventional time-series forecasting methods from the auto-regression family, namely, SARIMA and VAR. The results indicated that the individual-based model outperformed the auto-regression techniques by a significant margin, thanks to the deep neural network's uncanny ability in extracting useful features and capturing complex patterns within data. One of the major drawbacks of auto-regressions methods is their incompatibility with time-series that do not have equally spaced data points. Due to this fact, we were able to fit these models to the individual-level data to draw a direct comparison.

There are several limitations to highlight for this study. First, a reliable evaluation of the proposed management information system is only possible after implementing it and investigating its impacts on resource planning and quality of care productivity. This paper showed that the forecasting modules proposed could predict the demand with acceptable accuracy. In a comprehensive evaluation, the performance of the planning modules (shown in Fig. 3) and the effectiveness and efficiency of the entire system to achieve the goals in improving the quality of care and productivity should be assessed. Furthermore, the dataset used in this study is relatively limited. The lack of sufficient samples after cleaning the data certainly hindered the performance of the models and did not allow them to reach their full potential. Consequently, this work can serve as a proof-of-concept and can be improved upon with larger datasets with more extensive features as increasing the size of the training dataset would allow models to detect more complex patterns in patients' demands. Detected patterns in the training dataset may not be generalizable to other cancer control centers. This study aims to help other healthcare organizations replicate the proposed forecasting models and train them on their data.

6. Conclusion

Appropriate management of resources and workforce in healthcare systems is vital to maintain a nation in good health. Healthcare facilities such as cancer prevention and control centers that provide PC should coordinate their efforts to ensure that patients' needs are met adequately with appropriate quality. This study presents an informatics approach empowered by two LSTM-based neural network models built on the power of deep learning to forecast future demands of home hospitalized cancer patients at both individual and population-level. Experiments and results indicate that the presented models in this research are able to output reliable promising results, outperforming auto-regression techniques that are conventionally used for time-series forecasting, enabling managers to get a relatively clear image of the possibilities in the future.

Future works can consider including prediction of patients' families' needs where their quality of life is profoundly affected by the challenges of the illness. Palliative care considers the patient and the family as the focus of care and includes the bereavement period [50]. Another direction for future research is to design mathematical programming models that can address the discussed optimization tasks in our proposed framework.

This research stream would be helpful to provide effective automated data-driven decision making systems for home-based healthcare where the aging of the populations, and the growth of the very oldest segments within it, as a global trend, associated with the need for a transition from hospital-based to home-based healthcare delivery.

This study may also contribute to facilitating the broader use of machine learning in healthcare organizations in the form of a decision support system by demonstrating that the application of deep learning models can make a successful prediction of patients' needs. This approach could be applied in healthcare emergencies such as the current crisis with COVID-19 to make some important and helpful demand forecastings that will help healthcare executives.

CRediT authorship contribution statement

Marzieh Soltani: Literature review, Data analysis, Methodology, Interpretation of results, Writing - Original draft preparation. **Mohammad Farahmand:** Data analysis, Technical modeling, Implementation of the algorithms, Writing - Original draft preparation. **Ahmad Reza Pourghaderi:** Conceptualization a and design of this study, Supervision of modeling, Interpretation of results, Writing - Original draft preparation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank Ala Cancer Prevention and Control Center (MACSA) for their kind support. The authors also gratefully acknowledge support from Homayoun Naji Esfahani, director (Isfahan branch), Mohammad Reza Sharbafchi(MD), deputy director (technical), and Suzanne Hojjat-Assari (MD), deputy director (education) of MACSA, for providing methodological and clinical support during research.

References

[1] C. Sepúlveda, A. Marlin, T. Yoshida, A. Ullrich, Palliative care: the world health organization's global perspective, J. Pain Symptom Manage. 24 (2) (2002) 91–96.

[2] WHO, "Strengthening of palliative care as a component of comprehensive care throughout the life course," Sixty-seventh World Health Assembly, WHA67.19, 2014. [Online]. Available: https://apps.who.int/gb/ebwha/pdf_files/WHA67/A67_R19-en.pdf.

[3] G.I. Ringdal, M.S. Jordhøy, S. Kaasa, Family satisfaction with end-of-life care for cancer patients in a cluster randomized trial, J. Pain Symptom Manage. 24 (1) (Jul 2002) 53–63.

[4] J.M. Teno, B.R. Clarridge, V. Casey, L.C. Welch, T. Wetle, R. Shield, V. Mor, Family Perspectives on End-of-Life Care at the Last Place of Care, J. Am. Med. Assoc. 291 (1) (2004) 88–93 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/14709580/.

[5] L. Peters, K. Sellick, Quality of life of cancer patients receiving inpatient and home-based palliative care, J. Adv. Nurs. 53 (5) (2006) 524–533 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16499673/.

[6] B. Gomes, N. Calanzani, V. Curiale, P. Mccrone, I.J. Higginson, "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers," jun 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23744578/.

[7] B. Gomes, N. Calanzani, J. Koffman, I.J. Higginson, Is dying in hospital better than home in incurable cancer and what factors influence this? A population-based study, BMC Med. 13 (1) (2015) [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26449231/.

[8] G. Miccinesi, E. Crocetti, P. Morino, M. Fallai, M. Piazza, V. Cavallini, E. Paci, Palliative home care reduces time spent in hospital wards: A population-based study in the Tuscany Region, Italy, Cancer Causes Control 14 (10) (2003) 971–977 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/14750536/.

[9] M. Costantini, I.J. Higginson, L. Boni, M.A. Orengo, E. Garrone, F. Henriquet, P. Bruzzi, Effect of a palliative home care team on hospital admissions among patients with advanced cancer, Palliat. Med. 17 (4) (2003) 315–321 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12822847/.

[10] M. Ahlner-Elmqvist, M.S. Jordhøy, M. Jannert, P. Fayers, S. Kaasa, Place of death: Hospital-based advanced home care versus conventional care, Palliat. Med. 18 (7) (2004) 585–593 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15540666/.

[11] A. Hewat, C. Veitch, W. Cairns, "A retrospective review of place of death of palliative care patients in regional north Queensland," pp. 41–47, 2007. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17169959/.

[12] A. Alonso-Babarro, J. Astray-Mochales, F. Domínguez-Berjón, R. Gènova-Maleras, E. Bruera, A. Díaz-Mayordomo, C. Centeno Cortes, The association between in-patient death, utilization of hospital resources and availability of palliative home care for cancer patients, Palliat. Med. 27 (1) (2013) 68–75 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/22492481/.

[13] M. Riolfi, A. Buja, C. Zanardo, C.F. Marangon, P. Manno, V. Baldo, Effectiveness of palliative home-care services in reducing hospital admissions and determinants of hospitalization for terminally ill patients followed up by a palliative home-care team: A retrospective cohort study, Palliat. Med. 28 (5) (2014) 403–411 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24367058/.

[14] H. Seow, K. Brazil, J. Sussman, J. Pereira, D. Marshall, P.C. Austin, A. Husain, J. Rangrej, L. Barbera, "Impact of community based, specialist palliative care teams on hospitalisations and emergency department visits late in life and hospital deaths: A Pooled analysis," BMJ (Online), vol. 348, jun 2014. [Online]. Available: http://www.bmj.com/.

[15] P. Kubor, A. Pourghaderi, M. Moeeni, R. Rezayatmand, PCN234 cost-effectiveness analysis of home-based palliative care for end-stage cancer patients, Value in Health 22 (2019) S481 [Online]. Available: http://www.valueinhealthjournal.com/article/S1098301519328074/fulltext http://www.valueinhealthjournal.com/article/S1098301519328074/abstract https://www.valueinhealthjournal.com/article/S1098-3015(19)32807-4/abstract.

[16] M.J. Côté, S.L. Tucker, Four methodologies to improve healthcare demand forecasting, Healthcare financial management: J. Healthcare Finan. Manage. Assoc. 55 (5) (May 2001) 54–58.

[17] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Comput. 9 (8) (1997) 1735–1780 [Online]. Available: doi: 10.1162/neco.1997.9.8.1735.

[18] S. Chauhan, L. Vig, "Anomaly detection in ecg time signals via deep long short-term memory networks," in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015, pp. 1–7.

[19] J. Kim, J. Kim, H.L. Thi Thu, H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in: 2016 International Conference on Platform Technology and Service (PlatCon), 2016, pp. 1–5.

[20] V.K.R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, Chaos, Solitons and Fractals 135 (2020) 109864.

[21] L.K. Mojen, "Palliative Care in Iran: The Past, the Present and the Future," Supportive & Palliative Care in Cancer, vol. 1, no. 1, sep 2017. [Online]. Available: https://journals.sbmu.ac.ir/spc/article/view/11929.

[22] K. Zarea, M. Rassouli, M. Hazrati, S. Molavynejad, S. Beiranvand, Comparison of the hospice palliative care delivery systems in Iran and selected countries, International Journal of Cancer Management 13 (6) (2020) 1–17 [Online]. Available: https://sites.kowsarpub.com/ijcm/articles/101635.html https://sites.kowsarpub.com/ijcm/articles/101635.html#abstract.

[23] H. Heydari, S. Hojjat-Assari, M. Almasian, P. Pirjani, Exploring health care providers' perceptions about home-based palliative care in terminally ill cancer patients, BMC Palliative Care 18 (1) (2019) 1–9 [Online]. Available: https://link.springer.com/articles/10.1186/s12904-019-0452-3 https://link.springer.com/article/10.1186/s12904-019-0452-3.

[24] F. Anderson, G.M. Downing, J. Hill, L. Casorso, N. Lerch, Palliative Performance Scale (PPS): A new tool, J. Palliat. Care 12 (1) (1996) 5–11 [Online]. Available: https://journals.sagepub.com/doi/10.1177/082585979601200102.

[25] F. Ho, F. Lau, M.G. Downing, M. Lesperance, A reliability and validity study of the Palliative Performance Scale, BMC Palliative Care 7 (1) (2008) 1–10 [Online]. Available: http://www.biomedcentral.com/1472-684X/7/10.

[26] J. Harrold, E. Rickerson, J.T. Carroll, J. McGrath, K. Morales, J. Kapo, D. Casarett, Is the Palliative Performance Scale a useful predictor of mortality in a heterogeneous hospice population? Journal of Palliative Medicine 8 (3) (2005) 503–509.

[27] P. Prompantakorn, C. Angkurawaranon, K. Pinyopornpanish, L. Chutarattanakul, C. Aramrat, C. Pateekhum, N. Dejkriengkraikul, Palliative Performance Scale and survival in patients with cancer and non-cancer diagnoses needing a palliative care

consultation: a retrospective cohort study, BMC Palliative Care 20 (1) (2021) 74 [Online]. Available: https://bmcpalliatcare.biomedcentral.com/articles/10.1186/s12904-021-00773-8.

[28] S.S. Jones, A. Thomas, R.S. Evans, S.J. Welch, P.J. Haug, G.L. Snow, Forecasting daily patient volumes in the emergency department, Acad. Emerg. Med. 15 (2) (2008) 159–170 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18275446/.

[29] S.S. Jones, R.S. Evans, T.L. Allen, A. Thomas, P.J. Haug, S.J. Welch, G.L. Snow, A multivariate time series approach to modeling and forecasting demand in the emergency department, J. Biomed. Inform. 42 (1) (2009) 123–139 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18571990/.

[30] R. Champion, L.D. Kinsman, G.A. Lee, K.A. Masman, E.A. May, T.M. Mills, M. D. Taylor, P.R. Thomas, R.J. Williams, "Forecasting emergency department presentations, Australian health review: a publication of the Australian Hospital Association 31 (1) (2007) 83–90 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17266491/.

[31] N.R. Hoot, L.J. LeBlanc, I. Jones, S.R. Levin, C. Zhou, C.S. Gadd, and D. Aronsky, "Forecasting Emergency Department Crowding: A Prospective, Real-time Evaluation," Journal of the American Medical Informatics Association, vol. 16, no. 3, pp. 338–345, May 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19261948/.

[32] Y. Sun, B.H. Heng, Y.T. Seow, E. Seow, Forecasting daily attendances at an emergency department to aid resource planning, BMC Emergency Medicine 9 (1) (2009) 1–9 [Online]. Available: http://www.biomedcentral.com/1471-227X/9/1.

[33] D. Scarborough and M.J. Somers, Neural networks in organizational research: Applying pattern recognition to the analysis of organizational behavior.American Psychological Association, 2007. [Online]. Available: https://www.apa.org/pubs/books/4316077.

[34] M. Sangi, K.T. Win, F. Shirvani, M.R. Namazi-Rad, N. Shukla, Applying a novel combination of techniques to develop a predictive model for diabetes complications, PLoS ONE 10 (4) (2015) e0121569 [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121569.

[35] T. Vivekanandan, N.C. Sriman Narayana Iyengar, Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease, Comput. Biol. Med. 90 (2017) 125–136 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28987988/.

[36] M. Xu, T.C. Wong, K.S. Chin, Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network, Decis. Support Syst. 54 (3) (Feb 2013) 1488–1498.

[37] O. Olatubosun, F. Olusoga, F. Abayomi, Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure, Brit. J. Math. Comput. Sci. 6 (5) (2015) 381–393 [Online]. Available: www.sciencedomain.org.

[38] N. Süt, M. Senocak, Assessment of the performances of multilayer perceptron neural networks in comparison with recurrent neural networks and two statistical methods for diagnosing coronary artery disease, Expert Systems 24 (3) (2007) 131–142 [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0394.2007.00425.x https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0394.2007.00425.x https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0394.2007.00425.x.

[39] Purwanto, C. Eswaran, and R. Logeswaran, "A dual hybrid forecasting model for support of decision making in healthcare management," Advances in Engineering Software, vol. 53, pp. 23–32, nov 2012.

[40] J. Durbin, S.J. Koopman, Time series analysis by state space methods, Oxford University Press, 2012.

[41] V. Nair and G.E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21–24, 2010, Haifa, Israel, J. Fürnkranz and T. Joachims, Eds. Omnipress, 2010, pp. 807–814. [Online]. Available: https://icml.cc/Conferences/2010/papers/432.pdf.

[42] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1d convolutional neural networks and applications: A survey, Mechanical systems and signal processing 151 (2021) 107398.

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.

[44] H. Tabani, A. Balasubramaniam, S. Marzban, E. Arani, and B. Zonooz, "Improving the efficiency of transformers for resource-constrained devices," arXiv preprint arXiv:2106.16006, 2021.

[45] K. Smagulova, A.P. James, A survey on lstm memristive neural network architectures and applications, The European Physical Journal Special Topics 228 (10) (2019) 2313–2324.

[46] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Tech. Rep. 56, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html.

[47] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in Proceedings of the 26th Annual International Conference on Machine Learning, ser. ICML '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 41–48. [Online]. Available: doi: 10.1145/1553374.1553380.

[48] W. Zaremba and I. Sutskever, "Learning to Execute," arXiv preprint arXiv:1410.4615, Oct 2014. [Online]. Available: http://arxiv.org/abs/1410.4615.

[49] D.P. Kingma and J.L. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR, Dec 2015. [Online]. Available: https://arxiv.org/abs/1412.6980v9.

[50] M. Aparicio, C. Centeno, J.M. Carrasco, A. Barbosa, and M. Arantzamendi, "What are families most grateful for after receiving palliative care? Content analysis of written documents received: A chance to improve the quality of care," BMC Palliative Care, vol. 16, no. 1, pp. 1–11, sep 2017. [Online]. Available: https://bmcpalliatcare.biomedcentral.com/articles/10.1186/s12904-017-0229-5.