ARTICLE IN PRESS

Microprocessors and Microsystems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Microprocessors and Microsystems

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

Medical information retrieval systems for e-Health care records using fuzzy based machine learning model

Arokia Jesu Prabhu L [a], Sudhakar Sengan [b,*], Kamalam G K [c], Vellingiri J [d], Jagadeesh Gopal [e], Priya Velayutham [f], Subramaniyaswamy V [g,*]

[a] Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore-641062, Tamil Nadu, India
[b] Department of Computer Science and Engineering, Sree Sakthi Engineering College, Coimbatore – 641104, Tamil Nadu, India
[c] Department of Information Technology, Kongu Engineering College, Perundurai-638060, Tamil Nadu, India
[d] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India
[e] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India
[f] Department of Computer Science and Engineering, Mahendra Institute of Technology, Namakkal-637503, Tamil Nadu, India
[g] School of Computing, SASTRA Deemed University, Thanjavur – 613401, Tamil Nadu, India

ARTICLE INFO

Keywords:
Health care
Information retrieval
Machine learning model
Decision support
Clinical treatments

ABSTRACT

As other sectors advance through the aid of cognitive computing, whereas the health care sector is still evolving, offering more advantages for all consumers. The growing complexities of healthcare are compounded by an aging population that contributes to underprivileged decision-making contributing to adverse impacts on the standard of treatment and raises the cost of treatment. Advances in this field, however, is hampered by numerous challenges that create a gap between the knowledge base and user queries, query inconsistencies, and user domain information set. In recent years, the rapid development with the use of machine learning and artificial intelligence for medical applications has already been shown, from diagnostic heart failure to 1-D cardiovascular beatings and automated finding using multi-dimensional clinical data. Consequently, smart decision support structures are required, which can enable clinicians to make more informed treatment decisions. An innovative solution is to harness increasing healthcare digitization that produces enormous volumes of clinical data contained in e-HCR and merge it with advanced ML software to improve clinical decision-making, thus extending the medication evidence base at the same time. Through this work, we are investigating new methodologies as well as digging at specific real-life technologies already being implemented in the medical sector and concentrating mainly on studying about accurate depictions of patients from e-HCR.

1. Introduction

Automated labelling of documents is an efficient way of categorizing the documentation under predefined thematic categories at the document stage [1]. Health records, in which the medical files are published primarily in natural language, have been deemed a valuable tool for addressing different health problems by presenting specific patient circumstances, the clinical logic thought method, and clinical inference, which is typically not obtainable from the other components of the e-Health Care Records (e-HCR). Predictive modelling using e-HCR [2] intended to improve the quality of care, curb unnecessary expenditure and expand clinical knowledge at the same time. As exciting as it seems, clinical data comes with a multitude of obstacles to data-science that hinder the successful application of predictive models. The difficulties relate to the representation of data, the temporal processing, and the effect of data bias, which in turn drives the work discussed in this article [3].

We live in the age of algorithms, by which the Machine Learning (ML) [4] frameworks have influenced several industries like engineering, transportation, and management. In the past few decades, ML provides state-of-the-art output in numerous fields, ex: machine vision, text analytics, voice recognition, etc. Such technology has been inseparable from our daily existence, leading to the widespread application of ML algorithms in specific domains [5].

* Corresponding authors.
E-mail addresses: arokiajeruprabhu@gmail.com (A.J.P. L), sudhasengan@gmail.com (S. Sengan), kamalamparames@gmail.com, kamalamparames@gmail.com (K. G K), vellingiri.j@vit.ac.in (V. J), gjagadeesh@vit.ac.in (J. Gopal), priya.saravanaraja@gmail.com (P. Velayutham), vsubramaniyaswamy@gmail.com (S. V).

https://doi.org/10.1016/j.micpro.2020.103344
Received 14 August 2020; Received in revised form 4 October 2020; Accepted 15 October 2020
Available online 17 October 2020
0141-9331/

Recently, ML techniques show many outstanding results in multi-purpose tasks like identification of human body organs through digital medial images, detection of lung nodules, classification of interstitial lung diseases, reconstruction of medical images, and segmentation of brain tumors, to name a few. Today [6], more than anything, semantic engineering is used in the health care sector. The relationship between human beings and artificial intelligence aimed at improving global health care is no longer a dream. Health services are present in abundance, although most of that is still underused, such as health tracking services, although Smartphone devices. The opportunity for innovative technologies to evolve utilizing untapped, developed data offers an age for taking unparalleled medical-domain leaps forward [7].

Additionally, the patients do not often have diagnostic details accessible. While handling this data, it not only creates the subject of cognitive computing problems and also makes it a considerable data concern because of the enormous volume of data that has to be handled. Later in 2030 [8], it has a volume of usable medical data that would increase per 89 days, and till 94.56% of the data would stay unstructured e-HCR data with no specific spatial framework. Rather it represents an unmatched combination of multiple forms of evidence describing universal scientific principles and information regarding the individual and the treatment system. The description and representation in e-HCR of these clinical principles and patient data is a cornerstone of developing prediction models by ML [9]. The e-HCR data may be typically divided into structured and unstructured data. In this paper, we concentrate on e-HCR data, which is standardized. Structured data means documented information about the patient using controlled vocabulary instead of free text, drawing, or sound (Fig. 1) [10].

1.1. Research objective

The work priorities are viewed from a clinical and data analysis viewpoint. In this article, we present a systematic review of current literature on ML models' protection and robustness with a particular emphasis on their implementations in healthcare systems [11]. We also illustrate growing problems and risk factors that stand in the way of rigorous deployment of ML models in healthcare applications. This paper offers alternative strategies for addressing these problems. In summary, the specific contributions from this paper are as follows [12].

- The likelihood that an individual to achieve an inevitable result at a certain period provided the evidence available at the point.
- Medical information in an e-HCR is represented as a series of visits, and each visit contains patient details and clinical data.
- The aim here is to capture the order of e-HCR patient visits and events and include this information when creating predictive models.
- We devise the predictive healthcare ML pipeline and define can weakness sources at each point.
- We highlight various challenges relating to general securities and privacy, as well as those arising with the adoption of ML models.

2. Related works

While accessing the data, it starts increasing to prevalent because of the lower data storage and also hosting costs for effective IR techniques [13] that have never become higher. Also, specific systems can be varying in requirements for each application. As mentioned later, a medical program has substantially different demands than a traditional financial or trading stock framework, owing to the environment that includes its terminology. There are three sub-units in a text-based IR system: questions, database representations, and coordinating algorithms [14]. Although the design of each IR program can differ, and some algorithm may also be unknown to the public to increase the protection or finance, the basic framework is seen in Fig. 2 [15].

We have motivated the use of ML on e-HCR in facilitating decision support and reducing cognitive overload on doctors. Given that, it is essential not to ignore the importance of a comprehensive EHR itself [16]. A fundamental change in healthcare has evolved with advancements in scientific education and guidelines: from treatment in a centralized unit to treatment through several units of varied but specialist skills. In medicine, sometimes, it denoted as fragmentation. Consequently, patient care is divided along with a multitude with a variety of facilities in computer systems, and the assimilation of the information into a single system faces numerous challenges, mainly from an organizational point of view.

This includes security and privacy concerns, a lack of acceptable standardized data formats, and the use by disparate vendors of proprietary knowledge, expensive interface fees, and more. By

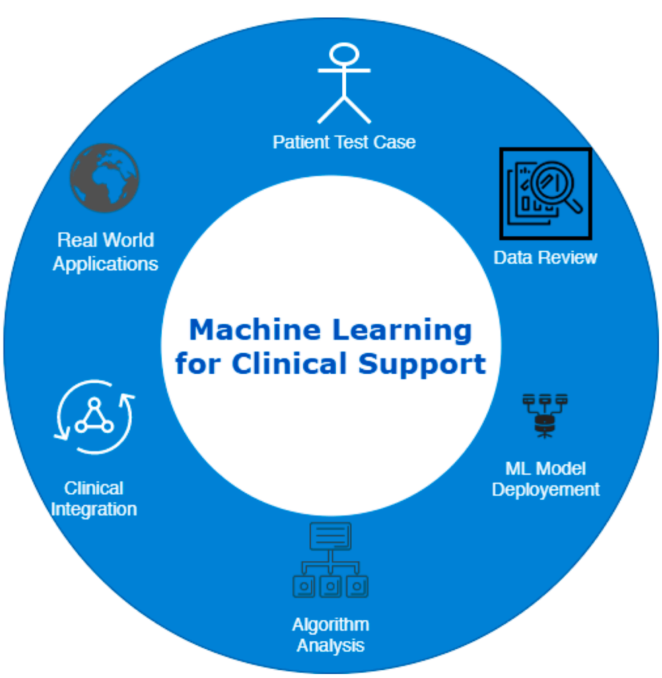

Fig. 1. The critical stages of ML's healthcare development for e-HCR.

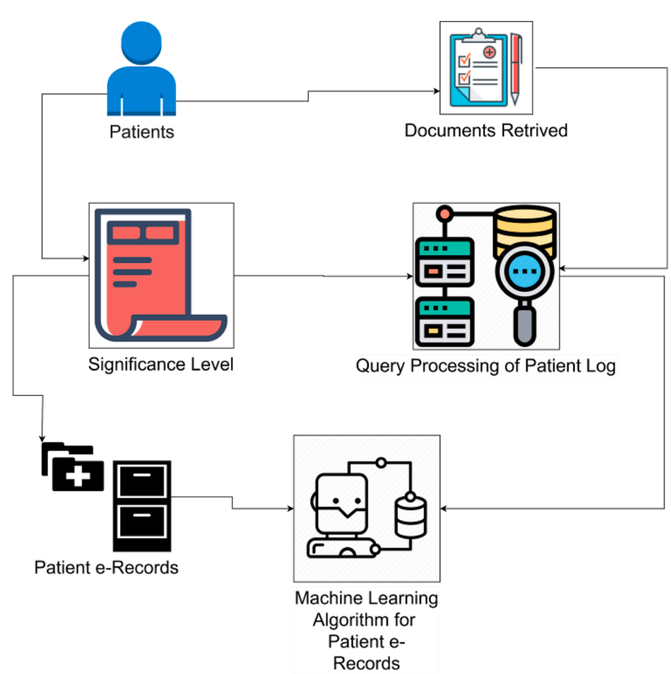

Fig. 2. Framework for ML-based IR.

consequence, service professionals have prohibited as receiving full data sets and are thus unable to grasp certain facets of the clinical experience of patients. It also contributes to repetitive clinical work and procedures, which add additional costs to the care system [17]. As with most e-HCRs, the data collection frequency in the platform depends on individual patient needs and is influenced by age and underlying morbidity. Consequently, the data represent real-world experience. Additionally, data is time-variant.

In a document, the data is stored in a particular file that contains an area of interest topics. A complete set of documents [18] is called a corpus. These types of documents are either automatically or manually transformed into a document representation to make it easier to match those with queries. Each representation of a document should reflect the intention of the author, where the context-based models and natural language processing becomes most important. If they are correctly indexed, a requirement of the user is balanced with representations of the text and returned to the user in a logical manner, such as listed off relevance [19]. Then the models are tested by calculating precision and accuracy for having their performance also analyzed. In precision, the percentage of documents obtained is necessarily considered for validating the percentage of valid documents when the process is obtained. Although these two factors usually negatively affect each other, both are also used to evaluate IR processes.

The suggested system is for recognizing people with type-2 Diabetes Mellitus using data from the Electronic Health Report [20]. In total, 300 patient samples took and about 114 attributes extracted by which various machine learning models were implemented like k-Nearest Neighbor, Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) [21], Naive Bayes and logistic regression. SVM provides the highest performance, with 96% precision from testing. Computer assistant classification is a system by integrating MRI convolution and MRI profusion data for the detection of various brain tumor forms and gliomal scoring. Samples were taken from 102 brain tumor patients, and they were added to endorse k-nearest neighbor, vector machine recursive feature exclusion, and linear classified study. The result exposed that SVM RFE provided the best outcome for tumor classification along with 85% accuracy and 88% for gliomal scoring. Defined various machine learning models, which include penalized are random forest models, logistic regression, and extreme gradient boosted decision trees primarily for high-risk surgical patient identification. They equipped Pythia system models, including electronic health reports with 194 clinical attributes including patient ages, smoking status, medications, co-morbidities, medical details, and surgical patient proxies.

Experimental results showed that the best result was obtained with the penalised linear regression model and an AUC value of 0.923. Five machine learning techniques are investigated penalized logistic regression, ANN with a single hidden layer, gradient boosting machine, random forest, and linear vector support machine for delirium risk forecast depends on e-healthcare data. A total of 18,333 trial patients had obtained, and studies were executed. It has been ascertained from the results where the gradient boosting algorithm formed the most acceptable outcome, along with an AUC value of 0.865. Proposed machine learning models to predict first-aid entry depends on EHR results. The authors implemented the COX with a dataset of 5.1 million patient observations to estimate the chance for first emergency admittance, and then it uses the algorithm for gradient boosting and random forest. They also established the gbm model by an AUC value of 0.779 worked best. The author used data from EHR to analyze the Seattle heart failure model for forecasting heart failure.

Samples were taken from 5044 patients, and characteristics were collected to determine the ranking for survival. The investigators first determined the survival score of Cox proportional regression patients that lived for I, II, or V years and then a patient who passes away later on V years was omitted. Several ML algorithms were applied to the remaining patients, including logistic regression, decision tree, random forest model, support vector regression, and a raise. Logistic regression worked better with a 10.16% change in the magnitude of the AUV curve from the experimental tests. Stefan H.

3. Current cognitive computing applications

3.1. EHR conversion

The migration of the e-HCR from conventional record storage methods is multiplying. Additionally, U.S. legislators approved Health Information Technology for Clinical and Economic Wellbeing Act that allows physicians to utilize EHR. Each technology used by health care professionals must be simple to use, as well as consistent with the relevant common nursing standards, as seen in Fig. 3.

A new paradigm is being developed for people with hearing impairments. Structured from three different modules and this paradigm offer a real-life implementation in the medical realm for cognitive computing. The first move is to obtain useful knowledge about safety from medical records. Next, a section on the knowledge base is used for identifying the details collected. The final module should store the information in a safe position that can be quickly reentered.

3.2. Clinical decision support

Clinicians can work with many specific cases, so they might even have a lot of medical details open for them. This is what clinician's responsibility for using the knowledge to identify patients accurately, and also to handle patients appropriately. However, to be able to utilize this material, physicians need to develop and retain a large body of expertise regarding illnesses, therapies, predicted results, and more. Substantial clinical trial studies and comprehensive evaluations help inform clinical practice, but the challenge is daunting and too essential to provide adequate medical treatment. Besides, e-HCR proliferation poses both

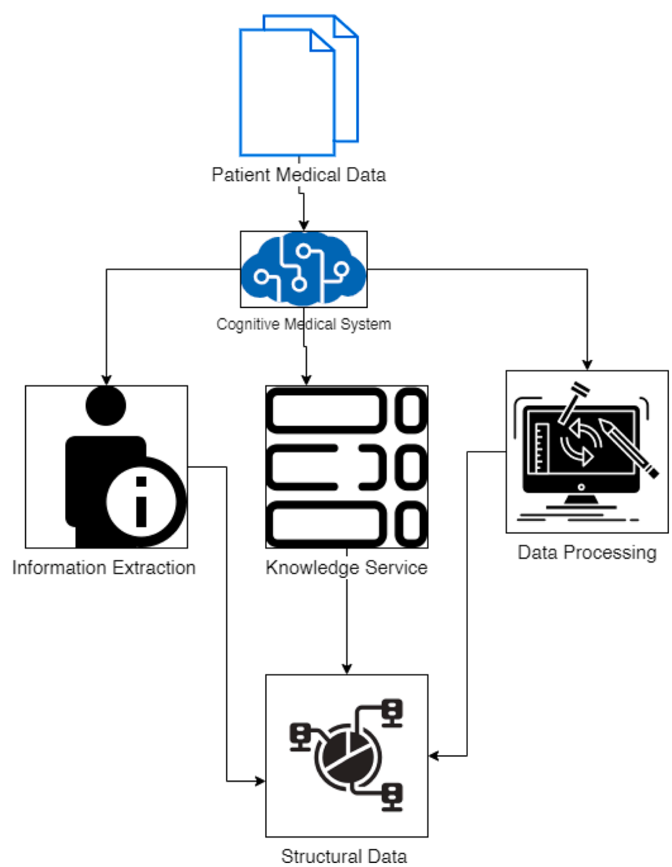

Fig. 3. Organizing of Patient Medical Log for ML.

gain to physicians and a strain on them. The downside is that knowledge may be collected for further usage in more standardized ways, but the disadvantage is that physicians have yet another ability set that must be mastered and preserved. This is perhaps shocking then that patient treatment through health care networks is complicated and sometimes suboptimal. The clinical judgement that allows patients to evaluate symptoms accurately and identify the disease and has significant help to improve patient treatment. Indeed, work is also shown the ability of these programs to support the clinical practice challenges, enhance clinician commitment for the procedure, and to strengthen overall health care systems. These techniques can come in various ways, but in clinical decision support, we concentrate on the can area of machine learning.

3.4. Clinical trials

Estimating the risk of the disease, which will be attributed to different treatments or exposures, is one of the fundamental challenges tackled in patient care and public health. Researchers perform clinical studies to study these reactions or therapies, finding randomized controlled trials as the gold norm for predicting Impacts of Medical Retrieval (IMR). In an RCT, patients in the sample are assigned to specific care methods, calculating the risk or likelihood of any result (Fig. 4). Randomization is critical because it eliminates conflicting factors, resulting in bias-free treatment impact measurements. The cumulative care impact is calculated by the disparity in result levels between diagnosis and regulation, as seen in EQU 1. Treatment is indicated either positive or negative IMR, along with the desirability of the calculated result deciding the success. The care arm with the highest rate of effectiveness is classified as a favored procedure. Although the IMR is representative of the real result of medication, though, we should predict a variety of results in humans. This makes levels of IMR less important to actual patients. The IMR based on population distribution, so generalizability to other test distributions is necessarily missing. Estimating the individualized results of medication, that provides the result for an individual's rather than the community level, would be much more useful (EQU (1)).

$$\mathrm{IMR} = \mathrm{P}(\mathrm{PatientTreatment} = \mathrm{True}) - \mathrm{P}(\mathrm{PatientTreatment} = \mathrm{False}) \tag{1}$$

It shows the presence of any impressive output (e.g., Cardiac Arrest), whereas others don't undergo with interest outcome. So that the monitoring group displays the calculated result rate of 57.1%, and the recovery community displays a rate of 28.6%. The IMR is, therefore, a lowering of the interest outcome rate.

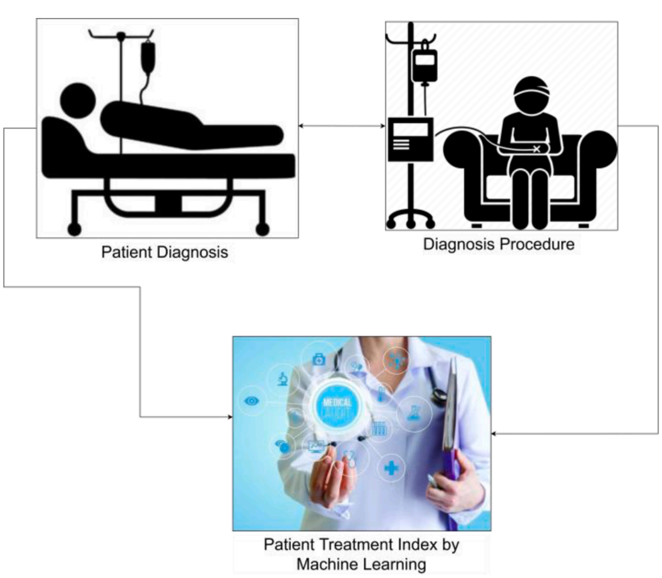

Fig. 4. Diagnosis of IMR for ML.

3.5. Decision tree model

The approach suggested by Papageorgiou notes that for the Fuzzy Cognitive Map Approach (FCMA), the concept of a methodology integrating decision tree along with a Decision Tree Algorithm (DTA) like ID3. To gain greater precision, the FCMA is equipped by using an unsupervised learning model [26]. The FCMA is equipped using the Hebbian learning algorithm. If suppose, a considerable quantity of the input data is given to the model, whereas quantifiable data is utilized for inducing decision trees and also for building FCMA. It makes Fuzzification of severe judgment authorizations that derived if-then statements, to increase the edibility. Using the pattern, a hospital in India reported 100 cases of bladder cancer. Using this,75 cases were labeled as low, and 30 labeled as high. After this evaluation, each tissue sample is retrospectively analyzed. FCMA rating was introduced with the use of a primary Bayesian classifier for output data to have distinct quantifiable values in both low and high-grade instances. In order to determine the efficiency, specificity, and sensitivity are determined for this model. The findings for the 95 samples were 79.18% sensitivity and 87.15% FCMA rating device specificity.

3.5.1. Algorithm of fuzzy based SVM model for ML based e-HCR

Input: SVM_I, S, F, SV, IV, B
SVM_I : Input Vectors
SVM_v : SNM Variables
SVM_F : Features of SVM
SVM_A→ Array in SVM
SVM_{AI}→ Array in Input SVM
SVM_B → Bias
f→function
D_t → Decision Tree
w→ Weights of the node
G → Best node of the decision tree
Output: O (Decision Function Output, Optimal Value for FSVM Classification)
Step 1: Start
Step 2: For i←1 to I by 1
Step 3: $O = 0$
Step 4: For j←1 to N by 1 Do
Step 5: D_t=0
Step 6: For k←1 to SVM_F by 1 Do
Step 7: $D_t \mathrel{+}= (SVM_v[j].f[k] - IV[i].f[k])^2$
Step 8: End
Step 9: G=Sort (S_1, S_2,S_3,….S_n)
Step 10: For $l = 1$ to M
Step 11: Input w
Step 13: Apply FSVM → To find newly generated solutions
Step 14: G= Best (Sort (S_1, S_2, S_3,….S_{n+m}), n)
Step 15: End
Step 16: K=ev(-¥ * D_t)
Step 17: O+=$SVM_V[j].\Phi$*k
Step 18: End
Step 19: $O = O + \Phi$
Step 20: End

4. Tools for data analysis

4.1. Data processing tools

Such properties may be presumed for the analysis of medical data to refine algorithms and enhancements in order to get optimal performance. First, data should be assumed to be broad in size, which is the case today for much of the data being utilized in multiple fields, purely because of the evidence accessible by technological advancement [4]. Secondly, details can be found to be inaccurate with specific attributes omitted due to the source details not providing such detail itself, as well as due to a lack of disclosure in health care providers' medical records due to regulatory limitations [9]. Additionally, the natural language of the data can be considered to be highly technical and include words that might not be widely documented to the general public. To achieve accurate outcomes, a knowledge-base can be integrated into the model. For example, data that includes the word "pyrexia", which is more generally referred to as "fever". Part of the task of developing a model involves understanding such subtleties.

4.2. Data set

We divided the data into two separate directions into the preparation and test sets. In case 1, we distinguish patient-based. 70% of patients used for train and 30% for the test, and on the test set, we recorded the results. We divide in case 2 depends upon the time. We used all the patient data for training from 2010 to 2015 and tested on the full 2015–2019 set. We test model results on visits that only took place in 2019, however. For the first example, we ensure no patient overlaps within the collection of preparation and study. So, the pattern is being checked on new patients that were not part of the curriculum. This is the standard approach in previous research on evaluating readmission predictions. However, predicting readmission danger for new patients and patients with previous admissions should be needed literally if a model used in a clinical function. The model efficiency of the above patients on subsequent visits may be enhanced if it includes their preliminary data in the exercise package. Thus, if the data are broken affording to case 2, we see improved outcomes (Fig. 5).

Phase I. **:** Create descriptions of visits from human and Machine-derived characteristics. For every visit, the output is a vector of apps.

Phase II. **:** The visit demonstrations are sequentially fed for processing in a cost-sensitive way.

Phase III. **:** The patients in the study are fed into the qualified network to assess the probability of 30 days of readmission at each visit.

We prove that by using both machine and human-derived e-HCR functionality together with LSTM network outstrips models which neglect both of this functionality. The model with these four features achieved an AUC of 0.77 (SD 0.006) and 0.82 (SD 0.003) for both cases 1 and 2 and called the most robust performance. Fig. 6 shows caseload results 1. We plan to explore the attention-based learning to have model intelligibility as the next stage of the project. Knowing what characteristics lead to a given result makes it possible to plan effective measures to

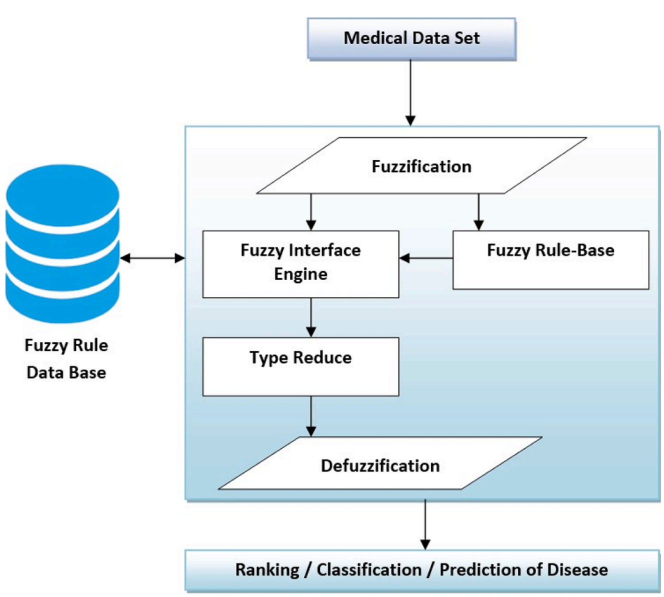

Fig. 5. Flowchart for Fuzzy Inferences on Medical Data.

mitigate or compensate for harmful outcomes. The model was conditioned and checked on 70.17% of patients on the remaining.

4.2.1. Decision tree

This review suggested the model CART (Classification and Regression Trees). Breiman et al. introduces it and is motivated to split the relative total number of squared errors into two partitions. Often, a two-step procedure is established, they are initial program induction by a train set under the theory method of "Divide and Conquer" and by the test set checking the procedure of accuracy. The hunt for CART splits on two covariate main characteristics for splitting on and the split point under covariate. First, decision trees are developed into a maximum stopping size when, due to lack of details, no splits can be produced again. Gini is close to the entropy test, which is used for grouping to break the law. By an II-decision target, the Gini measures node infection through the EQU (2),

$$GM(Treatment) = 1 - N(Treatment)^2 - (1 - N(Treatment))^2 \tag{2}$$

Where the comparative rate from one label on the node is expressed as ‘N(Tree).’ The tree is re-cropped to root depending on training data based on cost-complexity measurement EQU (3).

$$CC(Treatment) = CC(Treatment) + a|Treatment| \tag{3}$$

’R(T)’ is well-defined as the cost of training tree data set |T|, with ’T’ terminal node number, and then ’a’ specifies drawback forced on each node which increases from ’0′ to ’a’ value to be sufficiently pruned off for all splits. Consequently, and the next break is pruned to that tree’s overall efficiency.

4.2.2. Fuzzy inference system

Fuzzy Inference System (FIS) is a fuzzy logic set up by the mechanism precisely to render real-time judgments with human expert information in mind. The primary concept is intended to describe the inputs and outputs through a predefined series of rules that announces the heuristic information. Based on the theory of fuzzy sets, all variables (input or output) are defined with the linguistic variable ’DU_i’ where the value is described by linguistic values x_t^jthat fit into the discourse universe DU_i. The crisp values are unlike, and the universe values of the discourse are, to some degree, "belonging to" the linguistic value [0,1] described by a membership function $\mu(DU_i)$, and FIS_{iv}denotes the input value, EQU(4).

$$\mu_{x_i^j}(DU_i) = FIS_{iv} \rightarrow [0, 1] \tag{4}$$

The value ’1′ represents the full set value. Instead of a fuzzy value, the fuzzification process is needed to change the crisp values, and many methods are mostly used for singleton fuzzification. There, the assignment of i/p to o/p is defined by If-Else implication laws. To draw assumptions from i/p and the rule base, and inferential phase is necessary, EQU (5).

$$\begin{aligned} &IF \\ &DU_i \rightarrow x_i^1 DU_i \rightarrow x_i^2 DU_i \rightarrow x_i^3 DU_i, \ldots\ldots \rightarrow x_i^n \\ &THEN \\ &b_i = G_i() \\ &x_t^j \end{aligned} \tag{5}$$

In the above steps, "." gives the functional argument. The product of Takagi-Sugeno inference is a method that includes the terms data. At the end of the process, the defuzzification method is needed to get crisp output data: EQU (6)

$$y = \frac{\sum_{i=1}^{R} b_i \mu_i}{\sum_{i=1}^{R} \mu_i} \tag{6}$$

4.2.3. Fuzzification rules for decision tree

Usually, when the medicine is decided, and there will be no

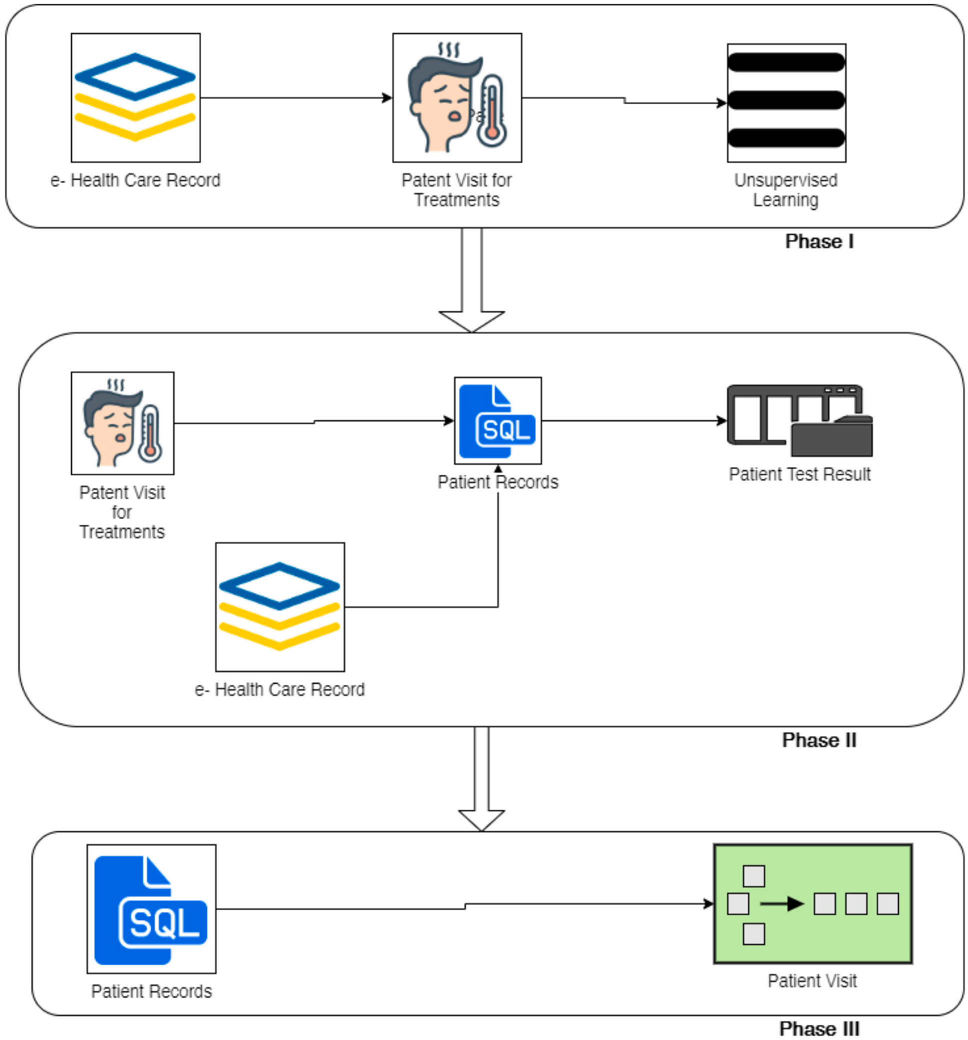

Fig. 6. Proposed framework.

predefined universal measure then. It is based primarily on the various intra- and inter-variability characteristics in which the process is involved. It does not make some meaning by this finding to be called crisp worth as the exact limit for decision making. This is one of the critical explanations for the inclusion of FIS in this report. Where fuzzification is the process of representing the medical input data into the knowledge base representation. Additionally, FL has constructed mainly on "categories" functions that are simpler for clinicians to interpret with their data group with similar characteristics for decision-making processes.

One of the critical components in the structure of FIS is the description of membership functions and a rules base. And the method of non-deep heuristic knowledge is incredibly challenging. The decision tree approach is built to dynamically achieve this from the entire order to avoid such kind of issue. The membership boundaries can be measured by the situations inherent in the test modules and rules. The limits of the decision tree, however, are crisp values depending on the train data, and FIS require fuzzy values. Besides this, the amount of trains and check data is reduced, and it transforms into danger again. To normalize the model and take advantage of fuzzy techniques, trapezoidal and triangular membership functions were used respectively for the discourse universe's central and edge dividers. In addition, we suggested that the limits of increasing membership feature should be expanded by 10% in order to obtain an overlap and also to prevent problems related to data restriction by training phase. Where new restrictions have been proposed for increasing membership feature, as shown below, EQU (7):

$$NewLowerLimit = LowerLimit - \frac{UpperLimit - LowerLimit}{2} \bullet 0.1$$

$$NewUpperLimit = UpperLimit - \frac{UpperLimit - LowerLimit}{2} \bullet 0.1 \tag{7}$$

4.2.4. Evaluation of the FIS decision maker

K-fold cross-validation should be carried out to determine the efficiency of the suggested approach and the ability of the subsequent FIS for forecasting decision-making behaviour. As the actual sample is randomly divided into sub-samples of k equal size and each sub-sample is to be considered for validation data to test the FIS, while the outstanding sub-samples of k-1 have been used as train data. The steps are replicated for k times, which differ with the validation details, and the outcomes are summed to find a single proof. Where the different variables were determined for classification outcomes in analysis and efficiency. The accuracy reflects the percentage of the dataset that is accurately categorized by ML classifiers. The sensitivity and specificity are calculated for positive and negative reports which are classified appropriately or not. Precision is calculated between recalled instances for a fraction of specific instances where the recall is a fraction of the relevant instances that are recalled from complete relevant instances. Below are defined by the mathematical expressions for measuring the different steps EQU (8).

$$Accuracy = \frac{TurePositive + TrueNegative}{TurePositive + TrueNegative + FalseNegative + TrueNegative} \bullet 100\%$$

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$Specificity = \frac{TrueNegative}{TruePositive + FalsePositive}$$

$$\text{Precision} = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$\text{Recall} = \frac{TruePositive}{TruePositive + FalseNegative} \quad (8)$$

By two classes, the True Positive represents positive records that are categorized precisely through FIS, and the True Negatives are false reports which are labeled strictly through the classifier. Whereas, False Positive is adverse reports (Fig. 7) that are incorrectly labeled, where False Negative represents positive reports classified incorrectly by the FIS.

4.3. Handling dataset annotation

One natural strategy for increasing the performance of ML models is for acquiring more labeled train data. It requires medical experts and radiologists to spend their valuable time interpreting medical data by manual, such as signals, medical images, and reports [25]. The other essential thing is the creation of accurate test sets, which is assessed the ML model's success and reveal shortcomings models. The manual sample note into the corresponding categories is thus costly, time-consuming, and tidy operation [22]. To address this problem, computational techniques can be established, and strategy is in active learning that can be used for annotating unlabeled data samples.

Information through different sources would be weighed while doing annotations on particular clinical uses because there might be a shortage of bright, standardized labelling of single-source reports. Such that the clinical data is highly used for the diagnosis purpose to get higher accuracy in prediction [23] [24]. The essential feature of ML is and integrating multiple source data known as phenol typing, in healthcare. To improve the ability of data annotators, recurrent deep models and NLP techniques will be used for retrieving and to incorporate rich knowledge from unstructured clinical reports (Fig. 8).

4.4. ML and distributed data management

The data is created in a dispersed manner in healthcare environments, i.e., through various divisions inside a hospital and also by separate hospitals. This involves the successful sharing and management of distributed data for clinical tests, predominantly using ML models. By this, full training and validation datasets are believed to be publicly generated and conveniently usable for designing ML models. Therefore, the creation of methods for distributed data processing and ML is rapidly required.

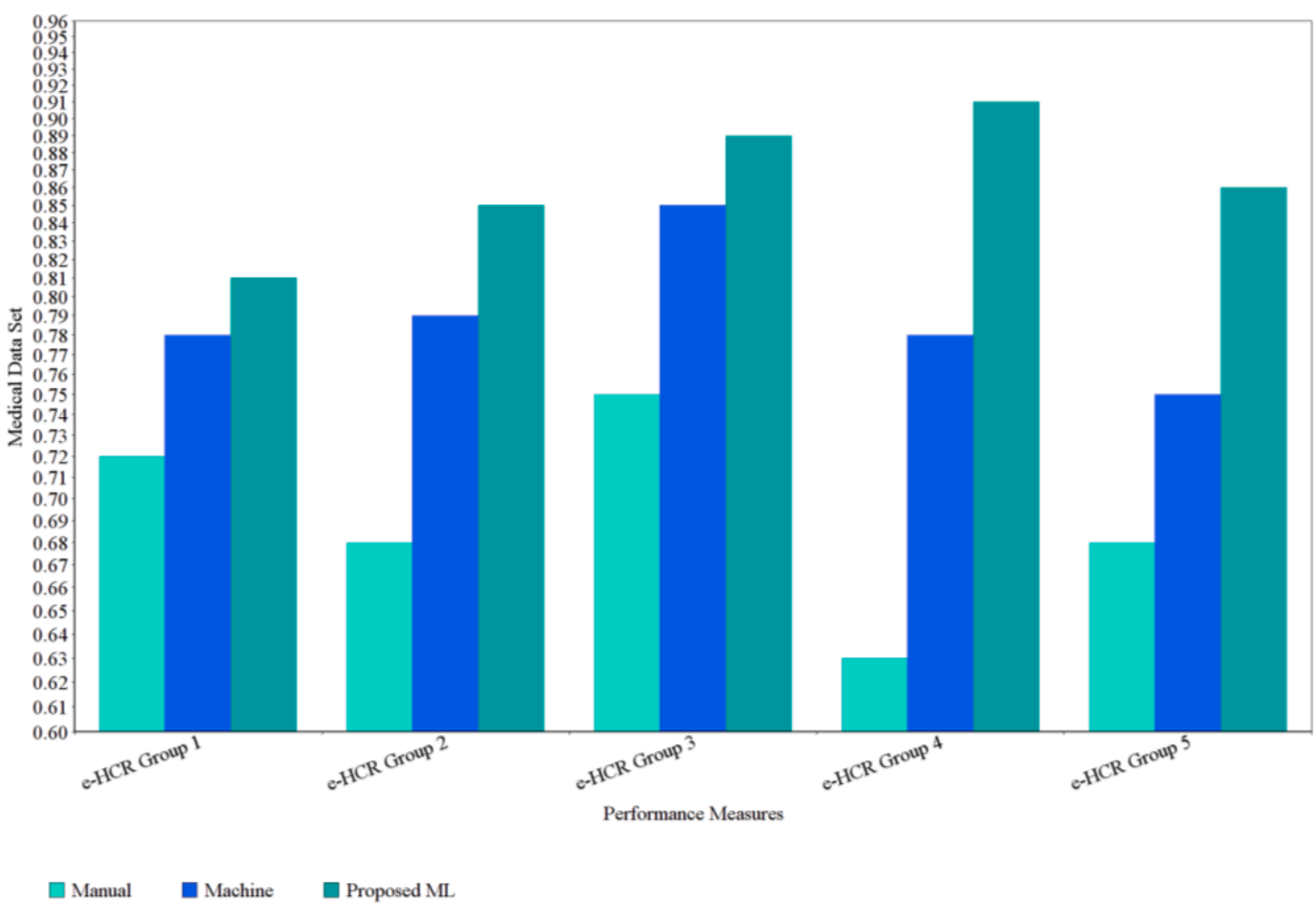

Fig. 7. The performance of the model in predicting the 30-day readmission of test patients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

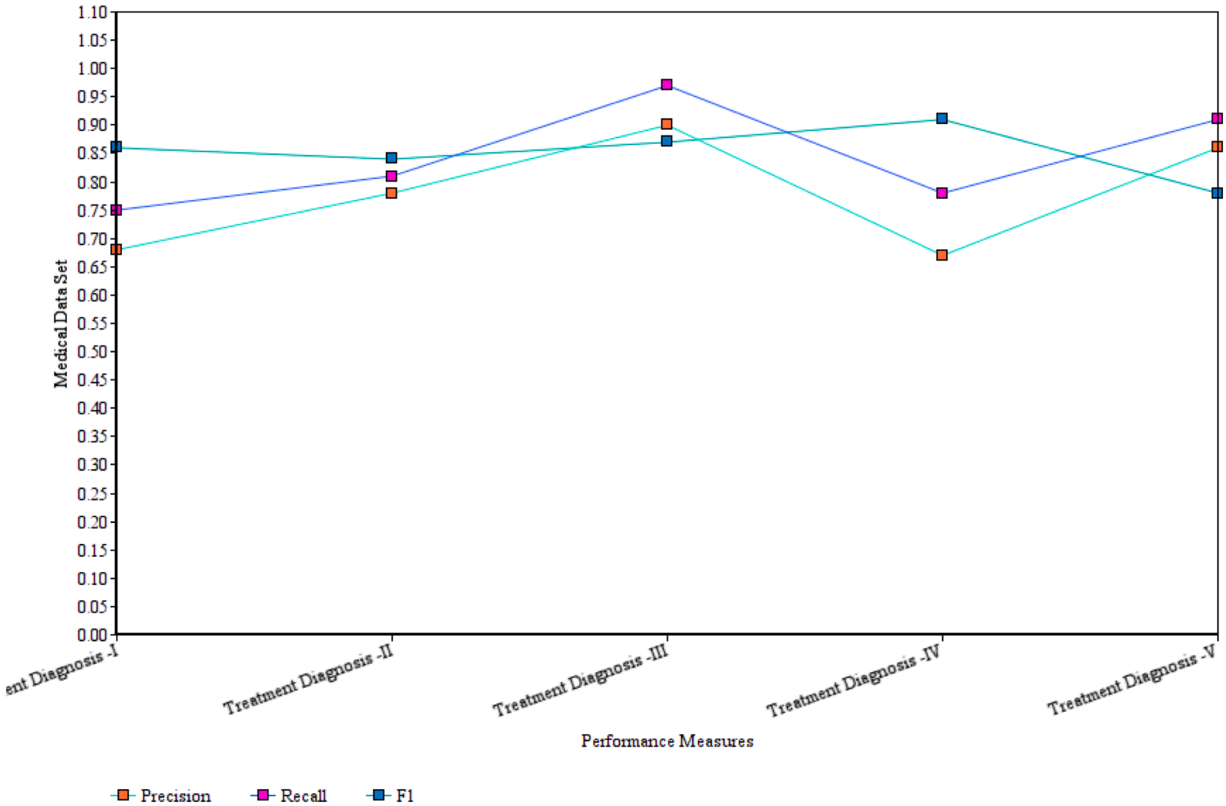

Fig. 8. An essential feature of ML.

4.5. Model-Driven ML

Although AI, ML, and big data are enormously valuable resources for e-healthcare, such technologies do not become panacea and necessary for mindful of related drawbacks and pitfalls. It fails to consider this, and one would quickly fall victim for the risky theory which data will and can talk for itself if accessible in abundance and support the development of theories as well — that in scientific terms which mean that the ML is adequate and irrespective of the needs for professional analysis, outer testing and comprehension of the provenance of evidence. It is necessary to integrate data-driven approaches with model-based or hypothesis-driven approaches and apply methodological rigour to such studies to prevent the numerous problems that emerge from inappropriate usage of ML in healthcare. Often, carefully constructed studies are required to extract causal explanations. Avenues for designing scientifically validated and reliable, safe, and effective ML approaches for healthcare need more community focus.

5. Conclusion and future work

In this article, we provided methods for portraying serial knowledge of patients in e-HCR to forecast adverse clinical outcomes. Appended to the relational embedding of scientific terms, the main issues of variability and temporality discussed utilizing human-derived characteristics. Use machine learning methods for clinical implementations has tremendous potential for improving the current provision of health care services. However, various privacy and security issues should be addressed to guarantee a robust and nonviolent application of the models in medical backgrounds. In this article, we delivered an outline of these problems by articulating the healthcare ML pipeline and finding growing points of weaknesses inside it. My research in promoting decision-taking has led by cooperation with healthcare professionals to apply machine learning models to specific clinical problems.

We have explored possible approaches for delivering privacy and secure-conserving ML for security-critical applications such as healthcare. Finally, we raised numerous exposed science concerns that would need additional study. This methodology will perform a significant role in future automation in health care, and it will produce more prediction and increase the accuracy level of the diagnosis process. This positively supports the medical practitioners, scientist, researchers etc.

Declaration of Competing Interest

There is '**No Conflict of Interest'** for submitting this article in your Journal for all authors.

References

[1] Nasser M Nasrabadi, Pattern recognition and machine learning, J. Electron. Imaging 16 (4) (2007), 049901.

[2] Geoffrey Hinton, Deep learning—A technology with the potential to transform health care, JAMA 320 (11) (2018) 1101–1102.

[3 Benjamin A Goldstein, Ann Marie Navar, Michael J Pencina, John Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, J. Am. Medical Inform. Assoc. 24 (1) (2017) 198–208.

[4 Nitesh V Chawla, Nathalie Japkowicz, Aleksander Kotcz, Special issue on learning from imbalanced data sets, ACM Sigkdd Explor. Newslett. 6 (1) (2004) 1–6.

[5] Robert A Verheij, Vasa Curcin, Brendan C Delaney, Mark M Mcgilchrist, Possible sources of bias in primary care electronic health record data use and reuse, J. Med. Internet Res. 20 (5) (2018).

[6 E. Punarselvam, Mohamed Yacin Sikkandar, Mohsen Bakouri, N.B. Prakash, T. Jayasankar, S. Sudhakar, Different loading condition and angle measurement of human lumbar spine MRI image using ANSYS, Springer-J. Ambient Intell. Humanized Comput., DOI 10.1007/s12652-020-01939-7,11, 2020.

[7 Menno Mostert, Annelien L Bredenoord, Monique Cih Biesaart, Johannes Jm Van Delden, Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach, Eur. J. Human Genet. 24 (7) (2016) 956.

[8 W. Duch, R. Adamczak, K. Grabczewski, A new methodology of extraction, optimization, and application of crisp and fuzzy logical rules, IEEE Trans. Neural Netw. 12 (2) (March 2001) 277–306.

[9 P. Olson, This AI just beat human doctors on a clinical exam," Jun 2018. [Online]. Available: https://www.forbes.com/sites/parmyolson/2018/06/28/ai-doctors-exambabylon- health/#2b20579b12c0.

[10] G. Tognola, A. Murri, D. Cuda, Cognitive computing for the automated extraction and meaningful use of health data in narrative medical notes: an application to the clinical management of hearing impaired aged patients, in: 2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), March 2018, pp. 299–302.
[11] G. Zuccon, B. Koopman, P. Bruza, Exploiting inference from semantic annotations for information retrieval: reflections from medical IR, in: Proceedings of the 7th International Workshop on Exploiting Semantic Annotations in Information Retrieval, ser. ESAIR '14, ACM, New York, NY, USA, 2014, pp. 43–45 [Online] Available, http://doi.acm.org/10.1145/2663712.2666197.
[12 P. Olson, This AI just beat human doctors on a clinical exam," Jun 2018. [Online]. Available: https://www.forbes.com/sites/parmyolson/2018/06/28/ai-doctors-exambabylon- health/#2b20579b12c0.
[13 F. Kuusisto, V Santos Costa, H. Nassif, E. Burnside, D. Page, J. Shavlik, Support vector machines for differential prediction, in: European Conference on Machine Learning (ECML-PKDD), 2014.
[14] H. Wang, Q. Zhang, J. Yuan, Semantically enhanced medical information retrieval systems: a tensor factorization based approach, IEEE Access 5 (2017) 7584–7593.
[15 Kawamoto, K., C. Houlihan, E. Balas, and D. Lobach. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 330, (7494):765.Peter Densen. Challenges and opportunities facing medical education. Transactions of the American Clinical and Climatological Association, 122:48, 2011.
[16 O. Ren, A.E. ohnson, E.P. Lehman, M. Komorowski, J. Aboab, F. Tang, Z. Shahn, D. Sow, R. Mark, L.W. Lehman, Predicting and understanding unexpected respiratory decompensation in critical care using sparse and heterogeneous clinical data, in: IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2018, pp. 144–151, 2018.
[17] Neveol Dalianis H., S. Velupillai, G. Savova, P. Zweigenbaum, Clinical natural language processing in languages other than English: opportunities and challenges, J. Biomed. Semantics 9 (1) (2018) 12.
[18 Arrhythmia Laboratory at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center). The MIT-BIH normal sinus rhythm database. Available online: https://physionet.org/physiobank/database/nsrdb/ (accessed on 21 January 2019).
[19 Denis Agniel, Isaac S Kohane, Griffin M Weber, Biases in electronic health record data due to processes within the healthcare system: a retrospective observational study, Bmj 361 (2018) k1479.
[20 Vijaya Kumar Veerabathiran, Devi Mani, Sangeetha Kuppusamy, Balu Subramaniam, Priya Velayutham, Sudhakar Sengan, Sujatha Krishnamoorthy, Improving secured ID-based authentication for cloud computing through novel hybrid fuzzy-based homomorphic proxy re-encryption, Springer-Soft Comput. (2020), https://doi.org/10.1007/s00500-020-05119-9.
[21] M. Kachuee, S. Fazeli, M. Sarrafzadeh, ECG heartbeat classification: a deep transferable representation, in: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics, New York, NY, USA, June 2018, pp. 443–444, 4–7.
[22 K. Rajakumari; M. Madhunisha, "Intelligent and convolutional-neural-network based smart hospital and patient scheduling system," IEEE, 2020 International Conference on Computer Communication and Informatics (ICCCI), DOI:10.1109/ICCCI48352.2020.9104173.
[23 Yang Jiao, Zhan Zhang, Ting Zhang, Wen Shi, Yan Zhu, Jie Hu, Qin Zhang, Development of an artificial intelligence diagnostic model based on dynamic uncertain causality graph for the differential diagnosis of dyspnea, Springer Link Front. Med. 14 (2020) 488–497.
[24] Jahanzaib Latif, Chuangbai Xiao, Shanshan Tu, Sadaqat Ur Rehman and Azhar Imran, "Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: a complete review, IEEE Access 8 (2020) 150489–150513.
[25] Sudhakar Sengan, Arokia Jesu Prabhu, L. Ramachandran, V. Priya, V. Ravi, Logesh Subramaniyaswamy, V., Images super-resolution by optimal deep AlexNet architecture for medical application: a novel DOCALN, IOS Press-J. Intell. Fuzzy Syst. (2020) 1–14, https://doi.org/10.3233/JIFS-189146.
[26 R. Vasanthi, R Jayavadivel, .K.. P.rasadh, J. Vellingiri, G. Akilarasu, S. Sudhakar, P. M. Balasubramaniam, A novel user interaction middleware component system for ubiquitous soft computing environment by using fuzzy agent computing system, Springer-J. Ambient Intell. Humanized Comput. (2020), https://doi.org/10.1007/s12652-020-01893-4.

L.Arokia Jesu Prabhu received his Bachelor degree from Anna University, Master degree from M.S.University and Pursuing Ph.D in Information and Communication Engineering from Anna University, Chennai. He has 11 years of teaching experience from reputed Engineering Colleges. His-research interests include Medical Image Processing and Cloud Computing. He has published 11 papers in reputed Conferences and Journals. CSTA and IAENG Membership holder.

Dr.Sudhakar Sengan is currently working as Professor in the Department of Computer Science and Engineering, Sree Sakthi Engineering College, Coimbatore, Tamil Nadu, India. He received Ph.D. degree in ICE from Anna University, Chennai, Tamil Nadu, and India. He has 20 years of Experience in Teaching/Research/Industry. He has published papers in 75 International Journals, 20 International Conferences and 10 National Conferences. He is a Research Supervisor in Anna University in Faculty of Information and Communication Engineering. His-research interest includes Network Security, Information Security and MANET, Cloud Computing, IoT. He received an award of Honorary Doctorate (Doctor of Letters-D. LITT.) from International Economics University; SAARC Countries in the field of Education and Students Empowerment in the month of April 2017. He has published Text book for Anna University syllabus: Title: "Digital Principles and System Design" Thakur Publications Pvt. Ltd, Chennai. ISBN: 978–93–87,880–77–1., Title: "Problem Solving & Python Programming", Charulatha Publication, Chennai. Title: Operating Systems", Thakur Publications Pvt. Ltd, Chennai. ISBN-978–93–88,809–15–3. He is a member of various professional bodies like MISTE, MIEEE, MIAENG, MIACSIT, MICST, and MIEDRC.

Dr.G.K.Kamalam is working as an Associate Professor in the Department of Information Technology, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India. She has a working experience of 22 years in teaching. She has completed her Ph.D in Anna University, Chennai. She is a recognized Supervisor in the Faculty of Information and Communication Engineering under Anna University, Chennai. Her research area is Grid and Cloud Computing. She has presented 35 papers in national and international conferences and published 32 research papers in reputed journals indexed by Scopus and SCI in her research and other technical areas. She authored three books on Design and Analysis of Algorithms, Test Your C Skills and Artificial Intelligence under Research India Publication, Vandhana Publication, and Sri Krishna Publication.

Dr. J. Vellingiri working as Assistant Professor (Senior), Department of Systems and Software Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India. He has a total of 14 years of experience in teaching and research. He has obtained his Doctoral of Philosophy (Ph.D.) from Anna University, Chennai in the year 2012. His-areas of research are Data Mining and Data Analytics. He has published several papers in national and international refereed journals and conferences. He is a life member of professional organizations such as Computer Society of India (CSI) and Indian Society for Technical Education (ISTE).

Dr. Jagadeesh Gopal is currently working as Associate Professor in Department of Systems and Software Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India. He has a total of 20 years of experience in teaching and research. He has obtained his Doctoral of Philosophy (Ph.D.) from Vellore Institute of Technology, Vellore, India. He obtained his Bachelors of Engineering (B.E) in Computer Science and Engineering and Master of Technology (M.Tech) in Information Technology. His-research interests include Software Engineering, Global Software Development, Soft Computing, Wireless networks. He has published several papers in national and international refereed journals and conferences. He is a member of various professional organizations such as Computer Society of India (CSI), Indian Society for Technical Education (ISTE), Indian Science Congress. He has reviewed various research papers.

Priya Velayutham has received her Ph.D. degree in Information and Communication Engineering in 2017 at Anna University, Chennai. Currently, she is working as an Associate Professor in Computer Science and Engineering at Mahendra Institute of Technology, Namakkal, Tamilnadu. She has more than 12 years of experience in Teaching and Research. Her research interests are in the areas of Artificial Intelligence, Machine Learning, Deep Learning, Cloud computing, Image processing, Data Science, Internet-of-Things and Bioinformatics. She published her research articles in reputed international journals, which is having a high impact factor. She received Best Faculty Award in Junior/ Department of CSE in 2019 from Shri P.K Das Memorial Best Faculty Award and Best Researcher Award in 2020 for her publications. She is a Life Member of the Indian Society for Technical Education (ISTE).

V.Subramaniyaswamy is currently working as an Associate Professor in the School of Computing, SASTRA Deemed University, India.. In total, he has 15 years of experience in academia. He has received the B.E. - CSE and M.Tech. - IT from Bharathidasan University, India and Sathyabama University, India respectively. He received his Ph.D. degree from Anna University, India and continuing the extension work with the support of Department of Science and Technology as a Young Scientist award holder. He has been contributing papers and chapters for many high-quality technology Journals and books that are being edited by internationally acclaimed professors and professionals. He is on the reviewer board of several international journals and has been a member of the program committee for several international/national conferences and workshops. He also serves as a guest editor for various special issues of reputed international journals. He is serving as a research supervisor, and he is also a visiting expert to various universities in India. His-technical competencies lie in recommender systems, cloud computing, Internet of Things, context-aware computing, Big Data Analytics, and Social Network Analysis.