International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

Computational Intelligence Based Electronic Healthcare Data Analytics Using Feature Selection with Classification by Deep Learning Architecture

Dr. S.A. Sivakumar
Associate Professor, Head-ECE,
Ashoka Women's Engineering College,
Kurnool - 518 218, Andhra Pradesh, India
drsasivakumar@gmail.com

Abstract:
EHRs (Electronic health records) are a source of big data that offer a wealth of clinical patient health data. However, because these notes are free-form texts, writing formats and styles range greatly amongst various records, text data from eHRs, such as discharge rapid notes, provide analysis challenges. This research proposed novel technique in electronic healthcare data analysis based on feature selection and classification utilizingDL methods. here the input is collected as input EH data, is processed for dimensionality reduction, noise removal. A public data pre-processing method for dealing with HD-EHR data is dimensionality reduction, which tries to minimize amount of EHR representational features while enhancing effectiveness of following data analysis, such as classification. The processed data features has been selected utilizingweighted curvature based feature selection with support vector machine. Then this selected deep features has been classified using sparse encoder transfer learning. the experimental analysis has been carried out for various EH datasets in terms of accuracy of 96%, precision of 92%, recall of 77%, F-1 score of 72%, MAP of 65%

Keywords: Electronic health records, feature selection, classification, deep learning, Dimensionality reduction

1. Introduction:

EHR systems maintain information about each person's health journey. Although the main purpose of EHR was to increase the effectiveness and accessibility of healthcare systems, it has also found extensive usage in clinical informatics and epidemiology. Early EHR analysis focused on less complex and better established statistical methods [1]. However, more recently, reliable prediction patterns in EHR data have also been discovered using statistical machine learning (ML) techniques including random forest, logistic regression, SVM, Cox proportional hazard model, and support vector machines (SVM). Even though these statistical models' ease of use and interpretability make them ideal for use in medical applications, their inability to handle high-dimensional input, reliance on numerous statistical and structural assumptions, and requirement for manually created features and markers make it impractical to use them for thorough analyses of EHR data [2].To address these problems, it is necessary to examine each person's complete medical background utilizing modelling tools that can find and take into account intricate nonlinear relationships between factors [3].

Although hospitals have successfully utilised the EHR for various administrative and corporate activities including patient logging, asset management, transfer management, and mostly billing operations, there is a need to discover strategies for successfully utilising the EHR for patient diagnosis [4]. The usage of EHR analytical solutions that will assist the doctor's knowledge is the sole remedy for this. Machine learning techniques, ranging from straightforward regression to intricate Recurrent Neural Networks (RNN), can now be employed to close the inferential gap for a variety of EHR tasks thanks to recent advances in artificial intelligence. The application of these learning systems to deliver effective care is hampered by a number of complicated integration challenges, the restricted availability of labelled data for training models, as well as privacy issues related to mistrust amongst providers. Even though deep learning methods are regarded as a cutting-edge innovation, there are still several EHR tasks that can be effectively performed using traditional machine learning methods including regression, random forests, and Bayesian methods. The most recent approaches in healthcare, such as computational phenotyping and the incorporation of genetic

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

data into therapeutic practises, have been made possible by machine learning [5].

Building health analytics solutions involves the laborious task of mining the longitudinal EHR data for clinical insights. Hospitals use specialised EHRs that are made up of a diverse range of components, many of which are large and unstructured content. Before deriving insights from the data, effective feature extraction and phenotyping are required due to the noise and sparsity of the EHRs. Although many studies have been conducted to examine the techniques used to extract data from electronic health records, it is important to comprehend EHR data mining from an aggregation point of view. For instance, a technique called adverse event prediction, which seeks to identify a hospitalised patient's approaching risk, can be carried out by combining information from sources such as medical records, MRI scans, ICD-10 nomenclature databases, and other sources. For this approach to gather insights from these divergent data, an analytical solution is required [6].

Contribution of this research is as follows:

- To propose novel technique in electronic healthcare data analysis based on feature selection as well as classification utilizing deep learning methods.
- To select the features of processed in data using weighted curvature based feature selection with support vector machine
- To classify deep features utilizing sparse encoder transfer learning

2. Background:

Utilizing a "well ordered and recorded database" of clinical data to learn from has a long history. Even though EHRs include a wealth of data, a recent systematic review of the medical literature discovered that predictive models created utilising EHR data only use a median of variables, rely on conventional generalised linear models, and use data from a single centre [7]. Most frequently used in clinical practise are simpler models, such as the 5-factor CURB-65 or single-parameter warning ratings. The lack of standards and semantic interoperability of health data from many locations has been a significant obstacle to utilising more of the data available for each patient [8]. Each new prediction task normally requires the selection of a distinct collection of variables, and the extraction and normalisation of data from many sources frequently necessitates a labor-intensive [9] procedure. Researchers have also reported contact-free sleep disorder detection utilizing sonar methods and built smartphone sensor-based applications for tracking sleep apnea [10]. The typical drawback of physiological monitoring methods is the need for additional intrusive monitoring equipment or professional oversight, which puts the alternate strategy of performing screening using routinely collected electronic health records front and centre [11]. As PSG is expensive, time-consuming, and labor-intensive, it is assumed that sleep physiological data likepulse oximetry as well as sleep stage length have significant prognostic capacity, but are not widely accessible [12,13]. Deep neural networks, which have many hidden layers instead of just one like ANN22–24, has wide range of applications in numerous fields [14,15]. Many deep learning neural network architectures, such as MLPs, RNNs, autoencoders, DBNs or CNNs, have been suggested [16]. In particular, CNNs have dominated the field of computer vision, and numerous variations are created over time. First model to incorporate convolution as well as pooling layers into a NN was known as LeNet-533, and their publication established the fundamental elements of CNNs [17]. However, it wasn't until 2012 that a CNN programme dubbed AlexNet34, which took first place in the picture classification, began to dominate the ImageNet 2012 competition. The CNNs' VGGNet, GoogleNet, and ResNet accomplishments are further noteworthy ones [18]. These methodologies investigate changes to the convolutional kernels and network architecture in an effort to reduce the size and increase the flexibility of CNNs while enhancing their performance [19]. It is demonstrated that CNN are adept at extracting local position-invariant characteristics from input for classification tasks and that CNN have found applicability to general NLP issues in recent years [20].

3. Materials and methods:

This section discuss novel technique in electronic healthcare data analysis based on feature selection and classification using deep learning techniques. here the input is collected as input EH data, has been processed for dimensionality reduction, noise removal. The processed data features has been selected using weighted curvature based feature selection with support vector machine. Then this selected deep features has been classified using sparse encoder transfer learning. the proposed architecture is shown in figure-1.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
Article Received: 10 June 2022 Revised: 20 August 2022 Accepted: 25 August 2022 Publication: 15 September 2022

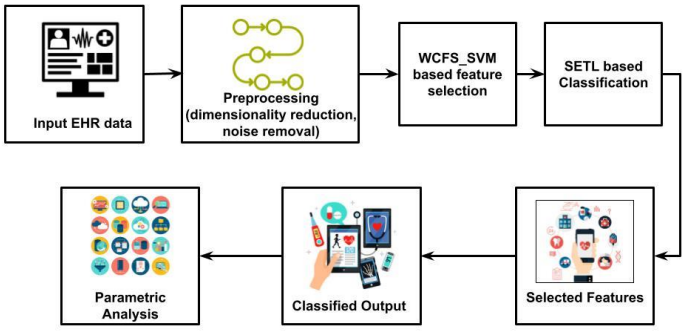

Figure 1. Overall proposed architecture

EHR data typically has a high number of dimensions and several input factors. It is important to remember that some input features might not be necessary for the solution to the issue at hand. Applying particular strategies to lower the dimensions of the original data set is a typical solution for dealing with such high-dimensional data successfully. In essence, feature extraction and feature selection are the two main components of dimensionality reduction strategies.

Weighted curvature based feature selection (WCFS) with support vector machine:

Reciprocal of radius of circle that permits between three points q1, q2, and q3 serves as a representation of the curvature. Only problems involving two-dimensional plane curves are taken into account in this paper. Given that the three points in a 2-D space are designated as q1(x1, y1), q2(x2, y2), and q3(x3, y3), and that E 2 and q1, q2, and q3 are non-collinear, MC on B is determined using equation (1):

$$\mathcal{MC}(q_1, q_2, q_3) = \frac{1}{R} = \frac{2\sin(\varphi)}{\|q_1, q_3\|} \qquad (1)$$

In this equation, R stands for the radius, ||q1, q3|| stands for the Euclidean distance between q1 and q3, and ϕ is angle of q2-corner of triangle formed by q1, q2, and q3, and it can be determined using the Law of Cosines equation (2):

$$\cos(\varphi) = \frac{\|q_1,q_2\|^2 + \|q_2,q_3\|^2 - \|q_1,q_3\|^2}{2\cdot\|q_1,q_2\|^2\cdot\|q_2,q_3\|^2} \qquad (2)$$

A data cleaning step and a data normalisation phase are present in the problem domain in real life. In this work, we use eq. (3) to normalise data using Min-Max (MM) method:

$$\mathcal{U}' = \frac{\mathcal{U} - \min(\mathcal{U})}{\max(\mathcal{U}) - \min(\mathcal{U})} \qquad (3)$$

This procedure ensures that our WCFS can fairly compare the curvatures for each characteristic and helps to cancel out the effects of any potentially significant big variances in the raw data set. Following is a description of the suggested WCFS method:

Step 1: 2D Data Re-construction: Cleaned HD data set U' is divided into n', 2-D planes as the first stage of the proposed CFS. This is accomplished by integrating all input attributes, $\mathcal{F}_i'(1 \leqslant i \leqslant n')$, and output y. As a result, U' is divided into n' 2-D planes, denoted by P(F 'i,y).

Step 2 – Feature Weighting: To find averaged curvature value of feature Fi' for every decomposed 2-D plane, the Menger Curvature method is used. Assumed that a deconstructed 2-D panel $\mathcal{P}_{(\mathcal{F}_i', y)}$consist of m data instances, Eq. can be used to get the Menger Curvature value $(\mathcal{MC}_{mj}^{i})$ of data point $\mathcal{F}_i'(2 \leqslant i \leqslant n')$. To this aim, eq. (4) is used to compute the mean of MC for F 0 I designated as $\widehat{MC}_{F_i'}$.

$$\widehat{MC}_{F_i'} = \frac{1}{2}\sum_{j=2}^{m-1} MC_{mj}^{i} \qquad (4)$$

where $(\mathcal{MC}_{mj}^{i})$ represents curvature value of m_j^{th} data point in feature F_i', $\widehat{MC}_{F_i'}$ indicates based on weight of feature F_i', greater value of F_i'.

Step 3 – Feature Ranking and Feature Selection: Based on the obtained $(\mathcal{MC}_{mj}^{i})$, the features are ranked using a traditional ordinal ranking approach. As a result, U 0's attributes are ranked. The necessary features are then chosen from the raw data collection U. The features that have $\widehat{MC}_{F_i'}$larger than the specified threshold ∂ will be chosen provided a threshold ∂. Equivalently, eq. (5) allows for the use of a TopK method:

$$u'' = U[Rank^{TopK}(\widehat{MC}_{F_1'})] \qquad (5)$$

in such a way that $\mathcal{U}' \in \mathbb{R}^{m\times n'}$ To achieve this, we kept the statistical properties of the initial data set while reducing the dimensionality of u to u".

Following division of |M| facets into G groups, determine mean Menger curvature for each group as well as the maximum and lowest Menger curvatures for each group. Assume that Kma,n mg and Kmean mg represent the maximum, minimum, and mean Menger curvatures of group mg, respectively. Mean Menger curvature Kmean mg of group mg is determined as indicated with |mg| the number of facets in the group mg by eq. It is average value of all Menger curvatures in mg (6).

$$K_{mean}^{mg} = \frac{\sum K_i \in m_g}{|m_g|} \qquad (6)$$

Next, as indicated in Equation, we describeΔmg as average of K mg max and K mg min (7). The specific value used to modify the mean Menger curvature K mg mean is indicated and written as Δmg.

$$\Delta_{mg} = \frac{K_{\min}^{m_g} + K_{\max}^{m_g}}{2} \qquad (7)$$

By altering the mean Menger curvature value of the group mg on basis of average value Δmg, each group mg is encoded in a watermark bit $\omega g \in \{0,1\}$ ($g \in [1, |G|]$). This implies that Kmean mg will change into Kmean mg that is

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

smaller than mg if $\omega g = 0$, $Kmeanmg$. If g = 1, Kmean mg will be changed by eq. (8) into Kmean$mg*$that is bigger than mg:

$$K_{\text{mean}}^{m_g^*} = \begin{cases} K_{mg*an}^{m_{m^*}} > \Delta_{mg} \text{ if } \omega_g = 1 \\ K_{\text{mean}}^{m_g^*} < \Delta_{mg} \text{ if } \omega_g = 0 \end{cases} \quad (8)$$

The watermarked mean curvature K mg mean is altered as stated in eqns (9) and (10) to satisfy the aforementioned embedding condition.

$$\omega_g = 1, K_{\text{mean}}^{mg*} = \begin{cases} \Delta_{mg} + \frac{\Delta_{mg} - K_{\text{mighan}}^{mg}}{m_m} \text{ if } K_{\text{mean}}^{mg} < \Delta_{mg} \\ K_{\text{mean}}^{mg} \text{ if } K_{\text{mean}}^{mg} > \Delta_{mg} \end{cases} \quad (9)$$

$$\omega_g = 0, K_{\text{mean}}^{mg*} = \begin{cases} \Delta_{mg} - \frac{K_{mix}^{mg} - K_{\text{minn}}^{mg}}{4} \text{ if } K_{\text{mean}}^{mg} > \Delta_{mg} \\ K_{\text{mean}}^{mg} \text{ if } K_{\text{mean}}^{mg} < \Delta_{mg} \end{cases} \quad (10)$$

the watermark bit ωg watermark to mean Menger curvature K mg∗ mean of group mg. The blue point designates mean Menger curvature, or K mg mean. Red point denotes the watermarked mean Menger curvature K mg∗ mean. If K mg mean is smaller than mg at ωg = 1, it will become bigger than mg. Reference value mg after embedding the watermark bit ωg into the group mg's mean Menger curvature, as indicated in Equation (11):

$$\alpha_g = \frac{K_{\text{mean}}^{mg*}}{\Delta_{mg}} \quad (11)$$

$$v'_{ij} = \alpha_g \times v_{ij} + (\alpha_g - 1) \times v_{ij} \forall j \in [1,3]$$

The objective is to establish a rule that will allow future explanations to be classified into the appropriate class using only the features that are currently available. One method for obtaining the rule is the Support Vector Machine (SVM). If the data can be linearly separated, SVM determines a boundary that divides two classes by maximising shortest distance between observations of each class and the border.s : Rp → Rl and seeking a linear discriminant function or a hyperplane by eq, (12) approach determines a nonlinear boundary when samples are not linearly separable

$$\beta^T \mathbf{s}(\mathbf{x}) + b = 0 \quad (12)$$

s(x) = (s1(x), . . . ,sl (x)) in feature space F, where β = (β1, β2, . . . , βl) is a l-dimensional vector of parameters. Aggregated margin between dividing boundaries can theoretically be maximised to get SVM's answer. The features that are used to build the rule should be sparse or limited in the interim to make it simple to put into practise. The SVM border is, mathematically speaking, the answer to minimising by eq (13),

$$Q(\beta, b, \xi) = \frac{1}{2} \| \boldsymbol{\beta} \|^2 + C \sum_{i=1}^{n} \xi_i \quad (13)$$

$$y_i(\beta^T \mathbf{s}(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{for } i = 1, \ldots, n \quad (14)$$

Equivalently, the Lagrangian dual function with form by eq. (15) can represent this optimization problem,

$$\underset{\kappa}{\text{Max}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j < \mathbf{s}(\mathbf{x}_i), \mathbf{s}(\mathbf{x}_j) > \quad (15)$$

subject to constraints by eq. (16)

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$
$$0 \leq \alpha_i \leq C \quad (16)$$

Inner product of two vectors xi and xj in dual function by eq is typically replaced by a scalar function K(·, ·) also known as a kernel function (17)

$$\text{Max}_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (17)$$

Kernel form of SVM boundary is expressed as eq(18) if we define SV as the set {j | αj > 0 for j = 1, 2, . . . , n} with all observations and xi, i∈ SV as SVMs.

$$\sum_{i \in SV} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b = 0 \quad (18)$$

and as a result, eq(19) is used to define the estimated bias term (bj) that was produced using the jth support vector (xj)

$$b_j = y_j - \sum_{i \in SV} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) \quad (19)$$

As a result, in real life, using the estimated coefficients of αi we may estimate b by taking the average of all estimated bjs with all support vectors. In practise, we may describe kernel function rather than locating projection mapping, even though kernel form of SVM. There are several frequently used kernel functions accessible, such as the radial kernel function via eq (20)

$$K(\mathbf{x}, \mathbf{z}) = h(-\| \mathbf{x} - \mathbf{z} \|^2) \quad (20)$$

The kernel function is then by eq(21) when h() originates from a Gaussian distribution with variance σ 2.

$$K(\mathbf{x}, \mathbf{z}) = \exp(-\| \mathbf{x} - \mathbf{z} \|^2 / 2\sigma^2) \quad (21)$$

A tiny change in x in input space, dx, translated into vector df in feature space so that by eq. (22), where f is mapped outcome of x ∈ R p in F.

$$\text{df} = \nabla \mathbf{s} \cdot \text{dx} = \sum_i \frac{d}{\partial x_j} \mathbf{s}(\mathbf{x}) \text{d}x_j$$

$$\nabla \mathbf{s} = \left(\frac{\partial (s(\mathbf{x})}{\partial x} \right) = \begin{pmatrix} \frac{\partial s_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial s_1(x)}{\partial x_p} \\ \vdots & \vdots & \vdots \\ \frac{\partial s_i(x)}{\partial x_1} & \cdots & \frac{\partial s_1(\mathbf{x})}{\partial x_p} \end{pmatrix} \quad (22)$$

As a result, the squared length of dfis expressed as eq (23) in quadratic form

$$\| \text{df} \|^2 = (\text{df})^T \cdot \text{df} = \left(\sum_i \frac{\partial}{\partial x_i} \mathbf{s}(\mathbf{x}) \text{d}x_i \right)^T \cdot \left(\sum_i \frac{\partial}{\partial x_j} \mathbf{s}(\mathbf{x}) \text{d}x_j \right)$$
$$= \sum_{ij} s_{ij}(\mathbf{x}) \text{d}x_i \text{d}x_j$$

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

$$s_{ij}(\mathbf{x}) = \left(\frac{\partial}{\partial x_i}\mathbf{s}(\mathbf{x})\right)^T \cdot \left(\frac{\partial}{\partial x_j}\mathbf{s}(\mathbf{x})\right) = \left(\frac{\partial s_1(\mathbf{x})}{\partial x_i} \ldots \frac{\partial s_j(\mathbf{x})}{\partial x_i}\right) \cdot \left(\frac{\partial s_1(\mathbf{x})}{\partial x_j} \ldots, \frac{\partial s_i(\mathbf{x})}{\partial x_j}\right)^T \quad (23)$$

As a result, the kernel K may be used to obtain the Riemannian metric, which is defined on the l × l matrix S(x) = [sij(x)], where S(x) is positive definite. Following lemma shows how a kernel function K and a mapping are related in a more simple manner.

In the traditional loss plus penalty form prediction risk framework, the classification boundary is evaluated by a linear weighted SVM by solving by eq(24), namely, Qi = c if yi = 1 and Qi = 1 c if yi = 1.

$$\min_{\boldsymbol{\beta},b} \text{Loss}\,(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta},\bar{b}} n^{-1}\sum_{i=1}^{n} Q_i\Big(1 - y_i(\mathbf{x}_i^T\boldsymbol{\beta} + b)\Big)_+ + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta} \quad (24)$$

where (1 t)+ = max (1 t, 0) is hinge loss, β are feature coefficients, b are intercept, and λ are positive regularisation parameters. Estimators of are given by eq (25) in an analytical form when hinge loss is viewed as E[Q(1 (yX T + b))+]

$$\tilde{\beta}_{\text{true}} = \arg\min_{b\beta} n^{-1}\sum_{i=1}^{n} Q_i\Big(1 - y_i(\mathbf{X}_i^T\beta + b)\Big)_+ \quad (25)$$

Additionally, imagine the true method contains sparse features, or, equivalently, that β T = (β T true, 0 T), where β T true = (β1, β2, . . . , βk). This will help you choose variables from the input space.A general form of penalty terms that would be directly added to loss function by eq. (26) was proposed in order to choose the vector z.

$$\text{Loss}\,(\beta) = n^{-1}\sum_{i=1}^{n} Q_i\Big(1 - y_i(\mathbf{x}_i^T\boldsymbol{\beta} + \beta_0)\Big)_+ + \sum_{j=1}^{p} p_\lambda(\|\beta_j\|) \quad (26)$$

where pλ(·) is a symmetric, non-convex penalty function with a tuning specification.

Allow f to represent mapped outcome of x ∈ Rp in F, that is, f = s(x) ∈Rl. The vector df in feature space will be mapped into a minor change in x in input space, dx, in such a way that by eq (27),

$$\mathrm{df} = \nabla\mathbf{s}\cdot\mathrm{dx} = \sum_j \frac{\partial}{\partial x_j}\mathbf{s}(\mathbf{x})\mathrm{d}x_{j,}$$

$$\nabla s = \left(\frac{\partial s(x)}{\partial x}\right) = \begin{pmatrix} \frac{\partial s_1(x)}{\partial x_1} & \cdots & \frac{\partial s_1(x)}{\partial x_F} \\ \vdots & \vdots & \vdots \\ \frac{\partial s_1(x)}{\partial x_1} & \cdots & \frac{\partial s_1(x)}{\partial x_F} \end{pmatrix} \quad (27)$$

As a result, the squared length of dfis expressed as eq in quadratic form (28)

$$\|\,\mathrm{df}\,\|^2 = \left(\sum_i \frac{\partial}{\partial x_i}\mathbf{s}(\mathbf{x})\mathrm{d}x_i\right)^T \cdot \left(\sum_i \frac{\partial}{\partial x_j}\mathbf{s}(\mathbf{x})\mathrm{d}x_j\right) = \sum_{ij} s_{ij}(\mathbf{x})\mathrm{d}x_i\mathrm{d}x_j \quad (28)$$

where the local magnification factor sij(x) can be thought of. Assuming C(x, x 0) is a positive scalar function, eq (29)

$$C(\mathbf{x}_i x') = c(\mathbf{x})c(\mathbf{x}') \quad (29)$$

where c(x) is a positive univariate scalar function and x and x 0 are feature vectors in input space. Kernel function K is then changed to eq (30)

$$R(\mathbf{x},\mathbf{x}') = C(\mathbf{x},\mathbf{x}')K(\mathbf{x},\mathbf{x}') = c(\mathbf{x})K\big(\mathbf{x},\mathbf{x}'\big)c(\mathbf{x}') \quad (30)$$

where K(x, x 0) is initial kernel function and K˜(x, x 0) is the second stage's updated kernel. One way to think about it is as a change of initial mapping s(x) to a new mapping function s˜(x), fulfilling eq (31)

$$s_{ij}(\mathbf{x}) = c_{ij}(\mathbf{x})s_{ij}(\mathbf{x}) \quad (31)$$

where sij(x) is described in (14) and $c_{ij}(\mathbf{x}) = \frac{\partial}{\partial x_i}\frac{\partial}{\partial \varepsilon_i}C(\mathbf{x},\mathbf{z})\Big|_{z=\mathbf{x}}$ This method is known as adaptive scaling, and it is simple to demonstrate that K satisfies Mercer positivity condition. When a suitable positive function c(x) is selected, the updated mapping s˜ can enhance separation.

Sparse encoder transfer learning based classification:

An unsupervised deep learning network called the autoencoder is utilized to minimize dimensionality of data as well as extract features.Fig. 2 illustrates a three-layer autoencoder that contains input layer, a hidden layer, and an output layer. By using a weight connection, original data are mapped onto concealed layer. The weight is fine-tuned to create accurate data representation while minimising reconstruction error. Due to the sparse limitation, certain buried layer nodes are active while the others are not, changing the autoencoder to SAE.

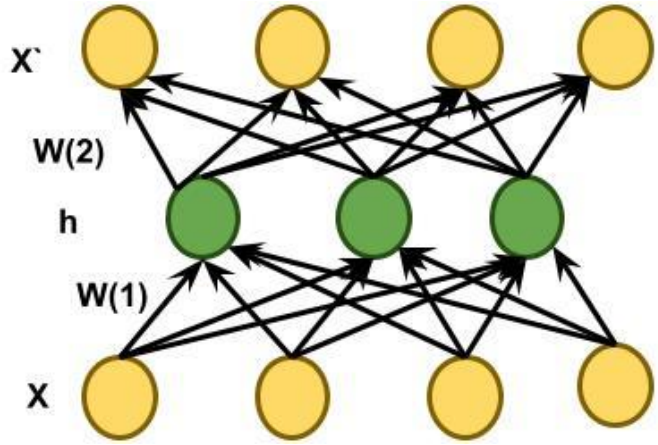

Figure. 2. Network architecture of autoencoder with three layers.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

The activation h of the hidden layer nodes is computed as eq. (32) for a sample x in the data set X = [x1, x2,...,xn] made up of n data samples

$$h = f\left(W^{(1)}x + b^{(1)}\right) \quad (32)$$

where b(1) is the bias and W(1) is weight used to connect input as well as hidden layer. In this study, the sigmoid function and the f activation function are both utilized. To rebuild original data by hidden layer as by eq. (33), connection weight between hidden and output layer is utilized.

$$\tilde{x} = f\left(W^{(2)}h + b^{(2)}\right) \quad (33)$$

where x˜ is data that has been rebuilt. Bias is defined as W(2), where W(2) is weight between hidden and output layer. With initial settings, a forward pass method is used to determine hidden layer's activation before reconstructing data in output layer. Reconstruction error is calculated for all data xi, I = 1,..., n in data collection in order to create an overall cost function of SAE network as eq (34)

$$J(W,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\|x_i - \tilde{x}_i\|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{n_i-1}\sum_{i=1}^{s_i}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^2 \quad (34)$$

where J(W, b) is the cost function that is being optimised for the two variables W and b. The parameters nl and l specify the number of network levels and the layer serial numbers, respectively. One hidden layer is utilised in network for nl = 3, while additional hidden layers are used if nl is more than 3. W(l) j I stands for all weight vectors connecting lth layer and l + 1st layer, whereas parameter sl stands for number of nodes in network's lth layer. Error in data reconstruction is the first component, and reducing this term can produce an accurate data representation. In order to prevent network form overfitting and restrict the weight's amplitude, the second term is a regularisation. The network weight and reconstruction error are adjusted using parameter λ. The mean activation of the jth node across data set is represented as eq in a hidden layer (35)

$$\rho_j = \frac{1}{n}\sum_{i=1}^{n} h_j(x_i), j = 1, \dots, s_l \quad (35)$$

with sl standing for number of nodes in fourth hidden layer. Activation of hidden layer is constrained by sparse parameter ρ, and the total constraint for every node in lth hidden layer is written as eq (36)

$$\sum_{j=1}^{s_j}\mathrm{KL}\left(\rho \parallel \rho_j\right) = \sum_{j=1}^{s_i}\rho\log\frac{\rho}{\rho_j} + (1-\rho)\log\frac{(1-\rho)}{(1-\rho_j)} \quad (36)$$

The KL(ρρj) is the Kullback-Leibler (KL) divergence, and this restriction is utilized to guide activation of nodes toward provided sparse sparse. Cost function of SAE is given as eq. when sparse constraint is taken into account (37)

$$J_{\text{sparse}}(W,b) = J(W,b) + \beta\sum_{j=1}^{s_f}\mathrm{KL}\left(\rho \parallel \rho_j\right) \quad (37)$$

where β and are two tuning parameters for the sparse penalty (W, b). The SAE network's cost function is computed following the forward pass processing. Back propagation approach is utilized to update initial weight as well as bias after solving the partial derivative of Jsparse to W and b.

Fig. 3 depicts a schematic representation of the deep SAE network with one input and three hidden layers. The kth SAE network, with k = 1, 2, and 3, is where the kth layer and (k + 1)th layer are found. (k + 1)th SAE receives its input from the hidden layer of the kth SAE.

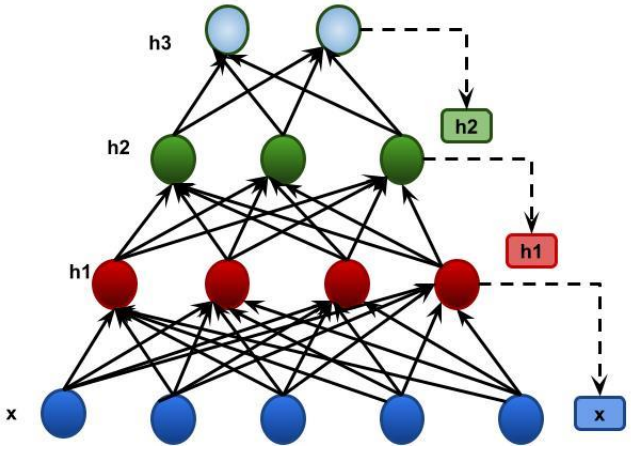

Figure. 3. Structure diagram of SAE network.

An unconstrained optimization issue is the minimization issue of Eq. (38) based on W1, b1, W2, b2, W0 2, b 0 2, W0 1, and b 0 1.We use gradient descent techniques to address this issue. For clarity, we first introduce following intermediate variables.

$$A_i^{(r)} = \left(\tilde{x}_i^{(r)} - x_i^{(r)}\right) \circ \dot{x}_i^{(r)} \circ \left(1 - \dot{x}_i^{(r)}\right).$$
$$B_i^{(r)} = \hat{\xi}_i^{(r)} \circ \left(1 - \hat{\xi}_i^{(r)}\right)$$
$$C_i^{(r)} = z_i^{(r)} \circ \left(1 - z_i^{(r)}\right). \quad (38)$$
$$D_i^{(r)} = \xi_i^{(r)} \circ \left(1 - \xi_i^{(r)}\right)$$

The objective Eq. (39)'s partial derivatives with regard to W1, b1, W2, b2, W 0 2, b 0 2, W 0 1, and b 0 1 is calculated, respectively, as follows:

$$\frac{\partial \mathcal{J}}{\partial W_1} = \sum_{i=1}^{n_e^*} 2\boldsymbol{W}_1'^{\top} A_i^{(*)} \circ \left(\boldsymbol{W}_2^{\top}\left(\boldsymbol{W}_2'^{\top} B_i^{(s)} \circ C_i^{(s)}\right)\right) \circ D_i^{(s)} \boldsymbol{x}_i^{(s)\top}$$
$$+\sum_{i=1}^{n_t^t} 2W_1'^{T} A_i^{(t)} \circ \left(W_2^{\top}\left(W_2'^{\top} B_i^{(t)} \circ C_i^{(t)}\right)\right) \circ D_i^{(t)} x_i^{(t)\top}$$
$$+\frac{\alpha}{n_s}\sum_{i=1}^{n_s} D_i^{(s)} o\left(1 - \frac{P_t}{P_s} + \ln\left(\frac{P_z}{P_t}\right)\right) x_i^{(s)}$$
$$+\frac{a}{n_t}\sum_{i=1}^{\mathrm{n_t}} D_i^{(t)} \cdot \left(1 - \frac{P_a}{P_t} + \ln\left(\frac{P_t}{P_s}\right)\right) x_i^{(t)T} + 2\gamma W_1$$

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

$$\frac{\partial \mathcal{J}}{\partial W_{2j}} = \sum_{i=1}^{n_s^*} 2W_{2j}'^{\top}\left(W_1'^{\top}A_i^{(*)} \circ B_i^{(v)}\right) \circ C_{ij}^{(s)}\xi_i^{(s)T}$$
$$+\sum_{i=1}^{nt} 2W_{2j}'^{T}\left(W_1'^{T}A_i^{(t)} \circ B_i^{(t)}\right) \circ C_{ij}^{(t)}\xi_i^{(t)\top} \quad (39)$$
$$\frac{\partial \mathcal{J}}{\partial W_2'} = \sum_{i=1}^{n_*} 2W_1'^{T}A_i^{(s)} \circ B_i^{(s)}z_i^{(s)T} + 2\gamma W_2'$$
$$+\sum_{i=1}^{n_t} 2W_1'^{T}A_i^{(t)} \circ B_i^{(t)}z_i^{(t)T}$$
$$\frac{\partial \mathcal{J}}{\partial W_1'} = \sum_{i=1}^{n,} 2A_i^{(s)}\xi_i^{(s)T} + \sum_{i=1}^{nt} 2A_i^{(t)}\xi_i^{(v)T} + 2\gamma W_{1,}^{t}$$

where nsj is number of instances with label j in the source domain, and W2j is j-th row of W2. We omit details due to space constraints because the partial derivatives of objective with regard to b1, b2, b 0 2 and b 0 1 are extremely similar to those of b1, b2, b 0 2 and b 0 1, respectively.

Algorithm of SAE_TL:

Input: Given one source domain $D_s = \left\{x_i^{(s)}, y_i^{(s)}\right\}\Big|_{i=1}^{n_s}$, and one target domain $D_t = \left\{x_i^{(t)}\right\}\Big|_{i=1}^{n_i}$, trade-off parameters $\alpha_{,}$ β, γ, number of nodes in embedding and label layer, k and c.

Output: Results of label layer z and embedded layer ξ.

1 StartW_1, W_2, W_2', W_1' and b_1, b_2, b_2', b_1' by Stacked Autoencoders performed on both source and target domains:
2 Evaluate partial derivatives of all variables
3 Iteratively update variables utilizing Eq. (18);
4 Redo Step2 and Step3 until methodmeets;
5 Evaluate embedding layer ξ and label layer z.

The process for classifying a new tool is as follows: the obtained monitoring data are first entered into new SAE network, weight transfer, weight update are then used, as illustrated in Fig. 4.

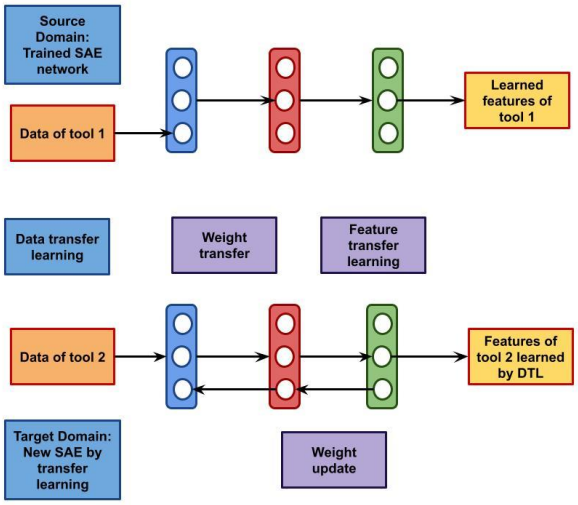

Figure 4. Architecture of SAE transfer network.

4. Experimental analysis:

Three open-source broker software solutions that were deployed using VMs were used to examine each of the distinct monitoring situations on version 3.1.Three VMs used to deploy broker software were hosted by Oracle, 2018 on a Windows 10 computer with 64GB RAM, an Intel Core i7-5820K processor, six physical CPUs, and twelve virtual CPUs operating at 3.30GHz.

Dataset description:

1)A project of the PCORI is PCORnet, National Patient-Centered Clinical Research Network. PCORnet is made up of a partner network that includes two HPRNs that are actively collaborating to link claims data with Electronic Health Records, 20 PPRNs, 13 CDRNs, and 13 PPRNs, all of which are based in healthcare systems like hospitals, integrated delivery systems, and EHR. EMR are standardised and stored uniformly by PCORnet using CDM.

2) GOS dataset:All information was retrieved from the Chinese Intracranial Hemorrhage Image Database retroactively (CICHID). From the hospital information system (HIS), all medical records and CT scans were exported and recoded for anonymization. Independent research assistants gathered the patient characteristics and scan parameters. Peking Union Medical College Hospital's Institutional Review Board gave its approval to the retrospective study (Ethics code:S-K1175). Each of the 2486 patients in the data collection has a single noncontrast computed tomography (NCCT) image and accompanying GOS. Patients with brain injury can objectively assess their recovery in five categories utilizing Glasgow Outcome Score. For ease of classification, we divide GOS into two groups: poor neurological results (GOS 3) and favourable

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

neurological outcomes (GOS > 3). 929 patients had successful neurological results, compared to 1557 patients who had unsuccessful neurological outcomes.

3) ADNI dataset: Data from Alzheimer's Disease Neuroimaging Initiative (ADNI) database were utilised to prepare this research. ADNI was established in 2003 under direction of Principal Investigator Dr. Michael W. Weiner. Main objective of ADNI is to determine whether serial MRI, PET, other biological markers, clinical, and neuropsychological evaluation may be used in conjunction to measure course of MCI and early AD.

Table-1 Comparative analysis of PCORnet dataset between proposed and existing technique

Parameters	VGGNet	GoogleNet	CI_EHDA_FS_DL
Accuracy	88	92	96
Precision	75	79	92
Recall	71	75	77
F1_Score	65	68	72
RMSE	75	73	68
MAP	72	68	65

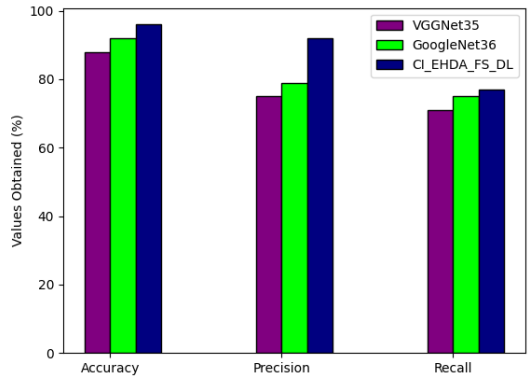

Figure-5 Comparative analysis of PCORnet dataset in terms of accuracy, precision, recall

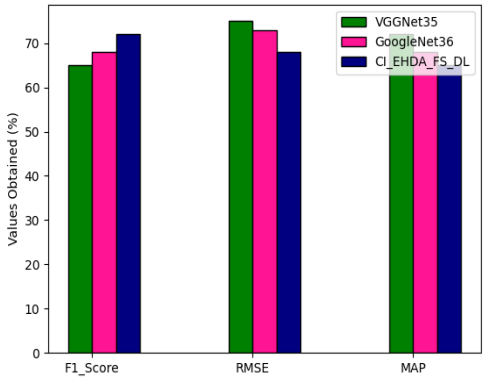

Figure-6 Comparative analysis of PCORnet dataset in terms of RMSE, F-1 score, MAP

Table- 2 Comparative analysis of GOS dataset between proposed and existing technique

Parameters	VGGNet35	GoogleNet36	CI_EHDA_FS_DL
Accuracy	82	86	92
Precision	75	78	83
Recall	79	83	85
F1_Score	82	85	88
RMSE	61	58	55
MAP	55	42	41

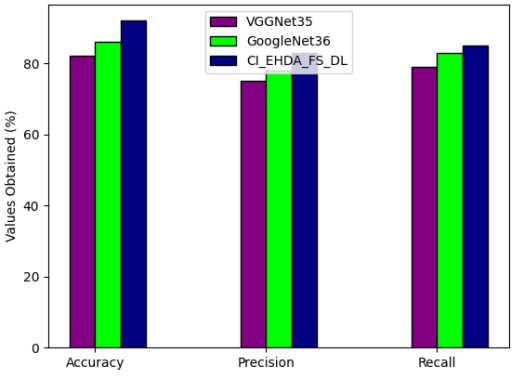

Figure-7 Comparative analysis of GOS dataset in terms of accuracy, precision, recall

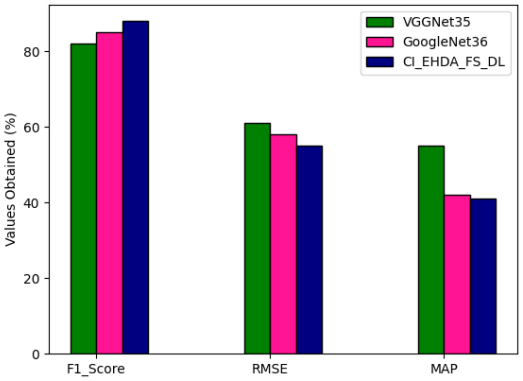

Figure-8 Comparative analysis of GOS dataset in terms of RMSE, F-1 score, MAP

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

Table- 3 Comparative analysis of ADNI dataset between proposed and existing technique

Parameters	VGGNet35	GoogleNet36	CI_EHDA_FS_DL
Accuracy	91	93	96
Precision	85	88	91
Recall	75	79	83
F1_Score	81	83	85
RMSE	81	75	71
MAP	75	71	65

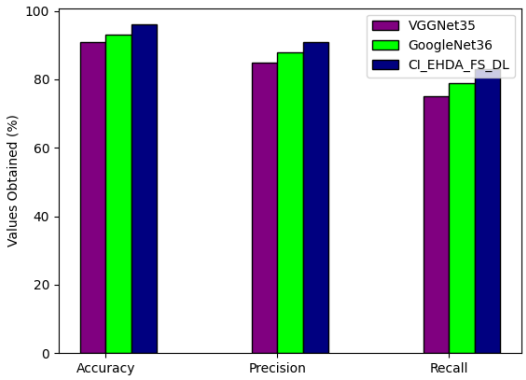

Figure-9 Comparative analysis of ADNI dataset in terms of accuracy, precision, recall

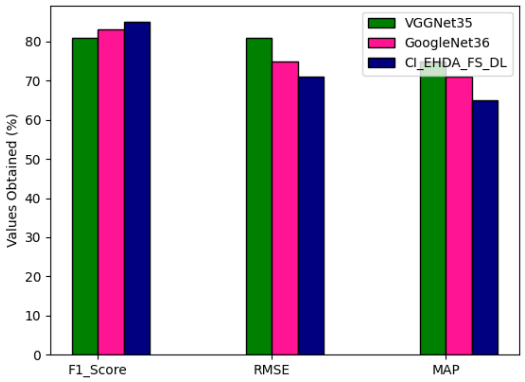

Figure-10 Comparative analysis of ADNI dataset in terms of RMSE, F-1 score, MAP

Above table 1-3 shows comparative analysis in terms of accuracy, precision, recall, F-1 score, RMSE and MAP. Here comparative analysis has been carried out based on various EH dataset like PCORnet, GOS, ADNI.Accuracy is one factor to consider when rating categorization methods. Accuracy is proportion of forecasts that our method successfully predicted. One indicator of the model's performance is precision, or the quality of a successful prediction. The total number of accurate positive predictions is divided by total number of real positives to determine precision. a model's ability to find all relevant instances in a data source. Calculating recall mathematically involves dividing the quantity of true positives by the total of true positives and false negatives.Precision: a classification model's capacity to isolate only the pertinent data points. The harmonic mean of recall and precision is used to calculate the F1 score. Recall that the harmonic mean is a substitute measure for the more often used arithmetic mean. It frequently comes in handy when calculating an average rate. We calculate the average of precision and recall for F1 score. Root mean square error, sometimes referred to as RMS deviation, is one of the techniques most frequently used to evaluate the accuracy of forecasts. It illustrates the Euclidean distance between measured true values and forecasts. The mean average precision (mAP), often known as average precision (AP), is a well-liked indicator for assessing how well models perform tasks like document/information retrieval and object detection.For PCORnet dataset proposed technique obtained accuracy of 96%, precision of 92%, recall of 77%, F-1 score of 72%, MAP of 65%. Where existing VGGNet obtained accuracy of 88%, precision of 75%, recall of 71%, F- score of 65%, RMSE of 75%, MAP of 72%. Googlenet attained accuracy of 92%, precision of 79%, recall of 75%, F- score of 68%, RMSE of 73%, MAP of 68%. Proposed method attained accuracy of 92%, precision of 83%, recall of 85%, F-1 score of 88%, MAP of 41%, VGGnet attained 82% of accuracy, 75% of precision, 79% of Recall, 82% of F-1 score, RMSE of 61%, 55% of MAP. GoogleNet attained accuracy of 86%, precision of 78%, recall of 83%, F- score of 85%, RMSE of 58%, MAP of 42% for GOS dataset. For PCORnet dataset proposed method obtained accuracy of 96%, precision of 91%, recall of 83%, F-1 score of 85%, RMSE of 71%, MAP of 65%. Where existing VGGNet obtained accuracy of 91%, precision of 85%, recall of 75%, F- score of 81%, RMSE of 81%, MAP of 75%. Googlenet attained accuracy of 93%, precision of 88%, recall of 79%, F- score of 83%, RMSE of 75%, MAP of 71% as shown figure 5-10.

5. Conclusion:

The proposed framework shows EH data analysis based on feature extraction and classification using deep learning

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

architectures. The input EH dataset feature extracted using weighted curvature based feature selection with support vector machine. Then this selected deep features has been classified using sparse encoder transfer learning. We uncovered the technological aspects of numerous initiatives that have been made to use deep learning models for clinical knowledge discovery using enormous data sets from electronic health records. Even while deep learning has been clearly successful for other hospital operations like billing and patient administration, there is still more to be done in the application of deep learning to EHR data. The proposed technique achieved accuracy of 96 percent, precision of 92 percent, recall of 77 percent, F-1 score of 72 percent, and MAP of 65 percent in the experimental study for varied datasets. Current methods addressed a number of key difficulties related to the depiction of EHR temporal data. In order to facilitate clinical adoption, future research concentrate on method transferability, clinical domain knowledge combination into research model, method interpretability improvement.

Reference:

[1]. Liu, X., Zhao, B., & He, W. (2020). Simultaneous Feature Selection and Classification for Data-Adaptive Kernel-Penalized SVM. *Mathematics*, *8*(10), 1846.

[2]. Zuo, Z., Li, J., Xu, H., & Al Moubayed, N. (2021). Curvature-based feature selection with application in classifying electronic health records. *Technological Forecasting and Social Change*, *173*, 121127.

[3]. Pham, G. N., Lee, S. H., Kwon, O. H., & Kwon, K. R. (2018). A watermarking method for 3D printing based on menger curvature and K-Mean clustering. *Symmetry*, *10*(4), 97.

[4]. Sun, C., Ma, M., Zhao, Z., Tian, S., Yan, R., & Chen, X. (2018). Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. *IEEE transactions on industrial informatics*, *15*(4), 2416-2425.

[5]. Zhuang, F., Cheng, X., Luo, P., Pan, S. J., & He, Q. (2015, June). Supervised representation learning: Transfer learning with deep autoencoders. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[6]. Sengan, S., Khalaf, O. I., Rao, G. R. K., Sharma, D. K., Amarendra, K., & Hamad, A. A. (2022). Security-aware routing on wireless communication for E-health records monitoring using machine learning. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, *11*(3), 1-10.

[7]. Haq, A. U., Li, J. P., Ahmad, S., Khan, S., Alshara, M. A., & Alotaibi, R. M. (2021). Diagnostic approach for accurate diagnosis of COVID-19 employing deep learning and transfer learning techniques through chest X-ray images clinical data in E-healthcare. *Sensors*, *21*(24), 8219.

[8]. Sengan, S., Khalaf, O. I., Sharma, D. K., & Hamad, A. A. (2022). Secured and privacy-based IDS for healthcare systems on E-medical data using machine learning approach. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, *11*(3), 1-11.

[9]. Zuo, Z., Li, J., Xu, H., & Al Moubayed, N. (2021). Curvature-based feature selection with application in classifying electronic health records. *Technological Forecasting and Social Change*, *173*, 121127.

[10]. Mitra, A., Rawat, B. P. S., McManus, D. D., & Yu, H. (2021). Relation classification for bleeding events from electronic health records using deep learning systems: an empirical study. *JMIR medical informatics*, *9*(7), e27527.

[11]. Garnica, O., Gómez, D., Ramos, V., Hidalgo, J. I., & Ruiz-Giardín, J. M. (2021). Diagnosing hospital bacteraemia in the framework of predictive, preventive and personalised medicine using electronic health records and machine learning classifiers. *EPMA Journal*, *12*(3), 365-381.

[12]. Ramesh, J., Keeran, N., Sagahyroon, A., &Aloul, F. (2021, October). Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning. In *Healthcare* (Vol. 9, No. 11, p. 1450). MDPI.

[13]. Han, S., Zhang, R. F., Shi, L., Richie, R., Liu, H., Tseng, A., ... &Tsui, F. R. (2022). Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *Journal of Biomedical Informatics*, *127*, 103984.

[14]. Si, Y., Du, J., Li, Z., Jiang, X., Miller, T., Wang, F., ... & Roberts, K. (2021). Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. *Journal of Biomedical Informatics*, *115*, 103671.

[15]. Zhao, Y., Fu, S., Bielinski, S. J., Decker, P. A., Chamberlain, A. M., Roger, V. L., ... & Larson, N. B. (2021). Natural language processing and machine learning for identifying incident stroke from electronic health records: algorithm development and validation. *Journal of medical Internet research*, *23*(3), e22951.

[16]. Venugopal, R., Shafqat, N., Venugopal, I., Tillbury, B. M. J., Stafford, H. D., &Bourazeri, A. (2022). Privacy preserving Generative Adversarial Networks to model Electronic Health Records. *Neural Networks*.

[17]. Lv, H., Yang, X., Wang, B., Wang, S., Du, X., Tan, Q., ... & Xia, Y. (2021). machine learning–driven models to predict prognostic outcomes in patients hospitalized with heart failure using electronic health records:

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 3
DOI: https://doi.org/10.17762/ijfrcsce.v8i3.2091
***Article Received**: 10 June 2022 **Revised**: 20 August 2022 **Accepted**: 25 August 2022 **Publication**: 15 September 2022*

retrospective study. *Journal of medical Internet research*, *23*(4), e24996.

[18]. Chi, C. Y., Ao, S., Winkler, A., Fu, K. C., Xu, J., Ho, Y. L., ... &Soltani, R. (2021). Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach. *Journal of medical Internet research*, *23*(9), e27798.

[19]. Ruiz, V. M., Goldsmith, M. P., Shi, L., Simpao, A. F., Gálvez, J. A., Naim, M. Y., ... &Tsui, F. R. (2022). Early prediction of clinical deterioration using data-driven machine-learning modeling of electronic health records. *The Journal of Thoracic and Cardiovascular Surgery*, *164*(1), 211-222.

[20]. Zhou, S. M., Lyons, R. A., Rahman, M. A., Holborow, A., & Brophy, S. (2022). Predicting Hospital Readmission for Campylobacteriosis from Electronic Health Records: A Machine Learning and Text Mining Perspective. *Journal of Personalized Medicine*, *12*(1), 86.