Artificial intelligence based health indicator extraction and disease symptoms identification using medical hypothesis models

L. Sathish Kumar1 · Sidheswar Routray2 · A. V. Prabu3 · S. Rajasoundaran1 · V. Pandimurugan4 · Amrit Mukherjee5 · Mohammed S. Al-Numay6

Received: 5 February 2022 / Revised: 14 July 2022 / Accepted: 28 July 2022 / Published online: 23 August 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Patient health record analysis models assist the medical field to understand the current stands and medical needs. Similarly, collecting and analyzing the disease features are the best practice for encouraging medical researchers to understand the research problems. Various research works evolve the way of medical data analysis schemes to know the actual challenges against the diseases. The computer-based diagnosis models and medical data analysis models are widely applied to have a better understanding of different diseases. Particularly, the field of medical electronics needs appropriate health indicator extraction models in near future. The existing medical schemes support baseline solutions but lack optimal hypothesis-based solutions. This work describes the optimal hypothesis model and Akin procedures for health record users, to aid health sectors in clinical decision-making on health indications. This work proposes Medical Hypothesis and Health Indicators Extraction from Electronic Medical Records (EMR) and International Classification of Diseases (ICD-10) patient examination database using the Akin Method and Friendship method. In this Health Indicators and Disease Symptoms Extraction (HIDSE), the evidence checking procedures find and collect all possible medical evidence from the existing patient examination report. Akin Method is making the hypothesis decision from count-based evidence principles. The health indicators extraction scheme extracts all relevant information based on the health indicators query and partial input. Similarly, the friendship method is used for making information associations between medical data attributes. This Akin-Friendship model helps to build hypothesis structures and trait-based feature extraction principles. This is called as Composite Akin Friendship Model (CAFM). This proposed model consists of various test cases for developing the medical hypothesis systems. On the other hand, it provides limited accuracy in disease classification. In this regard, the proposed HIDSE implements Deep Learning (DL) based Akin Friendship Method (DLAFM) for improving the accuracy of this medical hypothesis model. The proposed DLAFM, Convolutional Neural Networks (CNN) associated Legacy Prediction Model for Health Indicator (LPHI) is developed to tune the CAFM principles. The results show the proposed health indicator extraction scheme has 8–10% of better system performance than other existing techniques.

Keywords Artificial Intelligence · Medical Hypothesis · Deep Learning · Convolutional Neural Networks · Health Indicators

1 Introduction
A hypothesis is the method of finding the given input is right. Generally, raw data are stored in the database that is collected from various health sectors. Mostly, newly trained surgeons and physicians are an employee of the health sector to manage patient care systems. The proposed work has been executed utilizing the dataset ICD10 contains all disease categories along with disease type (Parkinson’s, heart issues, lung infections, cancer impacts, etc.), symptoms, location, and causes of disease. Each disease is denoted by an ICD10 code. The proposed hypothesis system will help and allow the user to get health indicators, which are diseases, symptoms, causes, and location based on the user’s input query. The most crucial medicines, the medical expert needs entitlements filers, obedience monitors, risk managers, and bookkeepers play a major role in a large healthcare organization. This system

Extended author information available on the last page of the article
ensures risk-free compliance with Medicare. The safe Medicare systems emerge under the acts such as the Affordable Care Act (ACA), Health Insurance Portability and Accountability Act (HIPAA), and CHIP Reauthorization Act (MACRA). The term prediction is used to classify the various types of symptoms of disease among the patient. The terms hypothesis are used to classify the disease that does not fall under any category.

It is used to find a solution under critical conditions. The evidence-based hypothesis is an effective method to give the correct solution to ambiguous medical problems. The existing medical techniques used mostly static decision-making rules on prescribed datasets. This leads to inaccurate disease detection rules. This practice has been considered a major research problem. In this regard, the proposed method has been implemented to carry out effective evidence-based medical hypothesis models to improve the system’s accuracy. The solution given by the evidence-based analysis is to be a better hypothesis for modeling any medical diagnosis system. As various existing schemes develop hypothesis models to build medical data extraction and diagnosis principles, the medical field gets computerized regularly. In this regard, the existing schemes are limited to analyzing and extracting the optimal health indicators from huge hypothesis models. This is a current research problem to be considered by the proposed system.

In the proposed method, the evidence checking model, Akin model, and medical hypothesis model handle the ambiguous health data indicators to support the patient health monitoring systems. On the other hand, the medical diagnosis report gives data category selection and traits-based extraction and friendship model for managing output records. In the evaluation phase, this proposed Health Indicators and Disease Symptoms Extraction (HIDSE) uses the ICD-10 database [1], EMR database [2], and World Health Organization database [3] for building intelligent medical hypotheses and health indicators finding schemes.

In the experiment base, the ICD-10 database has 43,000 records containing disease types (Parkinson’s, heart issues, lung infections, cancer impacts, etc.), symptoms, locations, causes, status codes, and disease descriptions. Similarly, the EMR database has 100,000 patient examination records. It contains patient identifiers, gender, age, admission identifiers, diseases, lab names, lab values, and lab units. Anyhow, the proposed system considers the health indicators such as tremors, stiffness, and voice disorders in Parkinson’s disease. Additionally, cancer-infected health indicators such as abnormal tissues, and growth factors are considered.

Finally, the WHO report has 20,000 statistics about deaths, lives, gender, year, etc. From the vast database entries, the proposed HIDSE technique extracts and classifies the health indicator using newly implemented medical hypothesis models.

The proposed HIDSE technique comprises medical data collection, Akin-based hypothesis model, trait-based feature extraction using the Friendship method, and DL-based CNN model for deep medical data analysis. In this case, the initial implementation setup has contributed to developing CAFM to extract the health indicators from huge medical datasets. This method has produced optimal results for existing models. At the same time, this proposed CAFM is limited to effective data reduction principles. This reduces the system’s accuracy frequently. In this condition, the proposed HIDSE framework contributed to implementing DLAFM layers using CNN training phases to increase the classification accuracy and precision of the system. This proposed DLAFM effectively handles the huge data features in the lower dimension by eliminating the useless data items gradually. This article contains a vast discussion on the various medical diagnosis and health indicator approaches in Sect. 2. Section 3 of this article describes the details of the proposed HIDSE using Akin and Friendship approaches, trait extractions, health indicator classifications, and technical algorithms. Section 4 shows the empirical observations and Sect. 5 concludes the proposed system.

2 Literature survey

There are different research works that contributed to useful medical findings. Notably, health indications and disease identifications are the major handlings of the following existing works. Feichen Shen et al. [4] proposed the method for identifying the risk factors of disease based on the bidirectional encoder representation, rule-based extraction from the precise medicine database. Authors collected the information in the format inheritance which properties derived from one class to another class method, and how the health indicators are traveled from one generation to another generation. The author also mentioned the Family Health Indicators (FHI) extraction based on the open health natural language processing and its Strengths, Weaknesses, Opportunities, and Threats analysis (SWOT) related to a healthy community also explained. However, this method failed to provide details about the common factors which was involved in the health indicators from one generation to another generation.

Kim et al. [5] developed the deep learning model for extracting the family medical health indicators which were highly precarious for affecting the organs. Natural Language Processing (NLP) has the prospective to extract the unstructured information to structure format to improve the patient health records for making us to better decision
making for the best diagnosis. In this model, the first portion consisted of family persons living status based on the health factors and the second portion consists of derived the common attributes of the super-entity of the data. Nevertheless, this model was nominal for entity relation and percentage of extracting health indicators for the decision-making systems \[6, 7\].

Baruch Vainshelboim et al. \[8\] found health hypotheses related to the Covid 19 situation important parameters which are related to the other diseases. Authors projected the health indication parameters which is the non-controversy of the face mask and it created the repository disease problem also increased acidity, toxify, short of breathiness, chronic stress all are the important parameters of the health extraction features of any communal disease. Emna Ammar et al. \[9\] presented the extracting the event logs related to health records for mining the health indicators content associated with instant data. Authors defined Business Process Management (BPM) as the key factor of health indicator continuous process improvement. Globally the patient was satisfied with the health indicators of the communal problem, which was used to improve the business process better than the other improving methods like clinical configuration, and clinical trials. Even though the indicators were not suitable for all the business process extraction methods of mining the data-centric.

Sam Henry et al. \[10\] mentioned the ranking system of health hypothesis indicators with the associated time slicing method of health indicators comparison to predict the extraction factor accurately. The authors evaluated four indirect measures which were direct co-occurrence of the ranking system of the non-controversy health indicator dataset.

This ranking system had the drawbacks of the low precision data set method of linking term-based occurrence, not to mention the health indicators properly. Frank S. Fan \[11\] proposed the hypothesis indicators for Parkinson’s disease, which provided protective effectiveness of diagnosing and lowering its incidence. Authors concentrated on the indicators of Parkinson’s disease may be the early stage and symptomatic stage of the disease. The effect of ractopamine was used for the protective effect of Parkinson’s disease, therefore great challenge and carefully designed investigations on this topic are necessary for the future. Zaydman et al. \[12\] proposed that epigenetic factors play a causal role in the health hypothesis.

The authors investigated scoliotic vertebral growth for the fracture indicators during vertebral fusion. Authors also acquire the phenomenon type of scoliotic deformities and the cells are the most powerful indicators of the chondrogenic and osteogenic cell differentiation. If the model is correct, then eliminating factors, and searching for factors were the difficult ones to find out the epigenetic factors. This method had problems in cell slicing datasets for identifying the epigenetic, health factors ostensibly.

Qiuli Qin et al. \[13\] proposed that medical record integrity is the important factor for all the consensus and non-consensus nodes of the proposed model. This model included data privacy, security, and integrity of the patient healthcare records. Any internal and external malicious nodes found in the model exposed the image of the blurred common shapes which was irrelevant to the extraction data. This model protected the tampering attacks of doctors and semi-activities of hospital health care persons. However, this model did not suitable for the smart contracts database and also the client functional database.

Pai et al. \[14\] provided information about the EMR latest innovations in healthcare Information Technology had modified the existing method of paperwork to be records. All medical records maintain manually is a highly hectic job for the healthcare information providers, so converting all the data into an electronic record is a possible and highly efficient way to treat the patient anywhere. Collecting various kinds of lab reports and medical images to store in an efficient manner clustering techniques may be used for reducing the storage space. The key focus of the hi-tech computerized act is to motivate the implementation of medical records and other supporting health systems in the world so that patients can get better care.

Julia Ostheimer et al. \[15\] developed Artificial Intelligence (AI) based medical record creation for the hybrid intelligent system, technological advancements of the efficient way of electronic medical records for patient and doctor centric view. During the development process, based on the important health care indicators iterative methods and agile methods are used to refine the execution model. Before starting the client–hospital designer model all the health hypothesis indicators and variants, and attributes should be stored in the data-centric for efficient decision making and the best way of using electronic records.

According to the current formulation of the better design-specific attributes to be used, that was the major drawback of the system. From the vast literature survey, this work identified the research problems and challenges. The existing systems worked on various medical supportive systems using vast amounts of practices. However, the available medical data extraction techniques were limited to identifying the disease symptoms using effective hypothesis/hypotheses models under uncertain health indications \[16–18\]. This proposed HIDSE system develops suitable health indicator analysis models and DL-based hypothesis structures for extracting the medical indications optimally.
3 HIDSE—the proposed system

The proposed HIDSE architecture has five phases, such as hypothesis deployment, trait selection, input collection, evidence checking, and an Akin-based hypothesis modeling system.

In the first phase, the hypothesis base selects all attributes of the EMR database to find the combination of all attributes and gives the hypothesis options. In the second phase, the scheme chooses the appropriate attributes for the hypothesis testing and then selects the respective attributes. In the third phase, this hypothesis model collects the medical data inputs of symptoms, suspect diseases, clinical names, and lab values. In the fourth phase, the collected input records are evaluated using the database features of ICD-10, WHO, and EMR. In this phase, the medical indicators of disease symptoms, and clinical measurements are checked and evidence reports are created for various patients [19].

In phase five, the Akin method analyses every attribute and clinical evidence and indicator count, to build the final hypothesis reports. The proposed HIDSE architecture has two schemes such as Akin Based health indicator extraction and Friendship based health indicator extraction. As indicated in Fig. 1, a novel hypothesis model works for health indicator extraction using Akin procedures. Similarly, Fig. 2 shows the Friendship based hypothesis model that works for health indicator extraction. Akin-based HIDSE has four phases, namely indicators input, category selection, traits extraction, and health indicators extraction. This method selects the traits for symptoms detection and checks the health evidence using EMR/ICD-10 databases. Finally, it produces hypothesis/hypotheses results.

On the other hand, the Friendship method gets the health indicators partial input, which is unstructured data. Every partial input comes under relevant categories to select the traits. Consequently, the features are compared with the WHO-Society of Exploration Geophysicists (WHO-SEG) database and EMR database where the categories are status, services, coverage, and risk factors. In the next phase, this scheme identifies category-based attributes from the existing patients’ EMR database and stores the observations in a common array. In the same way, it provides a hypothesis model that enables attribute selection and statistical observations. In this manner, this proposed HIDSE scheme develops medical hypothesis models, feature extraction models, and classification procedures to extract the medical data from complex EMR and ICD-10 databases [20–22]. The details are given in the following sections with appropriate technical benefits and algorithms based on Akin and Friendship procedures.

3.1 HIDSE hypothesis

3.1.1 Hypothesis setting

Procedure 1: Hypothesis input setting

Auxiliary Variables: Assign Symptoms as User Symptoms
Assign Disease as User Disease
Assign Lab Name as User Lab Name
Assign Lab Values as User Lab Values

Begin:

1. Possible Attributes: Symptom, Disease, Lab Name, Lab Values.
2. Hypothesis1: Symptom and Diseases
3. Hypothesis2: Symptoms, Clinical Test Name, Clinical Values
4. Hypothesis3: Diseases, Clinical Test Name, Clinical Values
5. Hypothesis 4: Symptoms, Disease, Clinical Test Name, Clinical Values

End: Hypothesis input setting

In this section, hypothesis setting takes the necessary EMR database for modeling the medical indicator hypothesis. It contains medical treatment and the examination values of the existing patients and database attributes. In this experiment, the EMR dataset contains symptoms, disease, and clinical outcomes. The examination of the correlation between various medical data attributes gives notable results in symptoms detection [23, 24]. Procedure 1 shows the hypothesis deployment rules.

Hypothesis 1: Symptom and Diseases
Hypothesis 2: Symptoms, Clinical Test Name, Clinical Values
Hypothesis 3: Diseases, Clinical Test Name, Clinical Values
Hypothesis 4: Symptoms, Disease, Clinical Test Name, Clinical Values

The inputs given in to the hypothesis model can be either normal data or suspicious data that to be recognized. The data are experimented with appropriate EMR, ICD-10 and WHO-SEG database features to give solutions [25]. Section 3.2.1 gives the medical indicators evidence check procedures.

3.1.2 Hypothesis evidence check procedure

Procedure 2: Propositions evidence check

Auxiliary Variables: Assign USYM as User Symptoms
Assign SYM, SCD as Symptoms array from ICD-10
Assign USDS as User Suspect Disease
Assign SR as Symptom Rank
Assign DIS as Diseases from ICD-10
Assign CD as code array
Assign FDIS, D as Final Disease array
Assign PL as Patient Lab Report Array
Assign LV as Lab value from the User
Assign SLN as User Select Lab Name
Assign SR, DR, DIR, LR as Counters and Rank Counters

Begin:
1. USYM and USDIS get from user
2. Get all SYM and SCD from ICD-10 Database
3. For SYM to SYM ((ii, i2 . . . in))
4. If USYM Contains (SYM(i)) Then
5. SR = SR + 1
6. CD(SR) = SCD(i)
7. End If
8. Next
9. Get all DIS and DCD from ICD-10 Database
10. If SR > 0 then
11. For DCD to DCD (i1, i2 . . . in)
12. If CD(i) Contains (DCD) Then
13. DR = DR + 1
14. FDIS(DR) = DIS (CD Code)
15. End If
16. Next
17. Get all DIS from ICD-10 Database
18. For DIS to DIS (i1, i2 . . . in)
19. If USDS Contains (DIS(i)) Then
20. DIR = DIR + 1
21. D(DIR) = USDS, End routine
22. Get PL from EMR Database based SLN
23. Select Which PL Result Disease matched with D(n)
24. For PL to PL (i1, i2 . . . in)
25. If PL(i) = LV or PL(i) Near By LV then
26. LR = LR + 1
27. End if

End: Propositions evidence check

For example, the data input deals with hypothesis 4, the following attributes shall be taken.

Symptoms: Rheumatoid myopathy with rheumatoid arthritis of left shoulder
Suspect Disease: Rheumatoid arthritis with rheumatoid factor
Clinical Test Name: CBC: Absolute Lymphocytes
Clinical Values: 31

This practice tests whether the user symptoms are the suspect disease symptoms and the clinical values are examined with the evidence from the EMR patient database. The user’s symptoms are also matched with ICD-10 symptoms and symptoms codes. Accordingly, all diseases are taken into examination bed using normal and suspect data features of ICD-10 and EMR features.

3.1.3 Akin procedures

The Akin procedure takes the evidences count for various input data and gives the result through fair hypothesis channel. The evidence and Akin value have been considered as final hypothesis metric.

\[
\text{Evidence Count} = \sum_{i=1}^{n} EVC_i / \text{Number of Evidence Count} \tag{1}
\]

Akin Method = \sum_{i=1}^{n} \text{Evidence Count} > 50\% \tag{2}

where, EVC—Evidence count. EVC1 – Cn is denoted disease and symptoms are matched with existing EMR.

If the evidence counts for the symptoms exceeds more than five record matching for all attributes of the database, the overall evidence count result reaches 50% that is
assured through the Akin method. Procedure 3 shows the Akin procedures.

**Procedure 3: Akin method**

Auxiliary Variables: Assign SR, DR, DIR, LR as Counters and Rank Counters

Begin:

1. If SR > 5 and DIR > 5 and LR > 5 then
2. Result: Hypothesis 4 is Conform
3. Else if SR > 5 and DIR > 5 then
4. Result: Hypothesis 1 is Conform
5. Else if SR > 5 and LR > 5 then
6. Result: Hypothesis 2 is Conform
7. Else if DIR > 5 and LR > 5 then
8. Result: Hypothesis 3 is Conform
9. Else
10. Result: Not Possible for this Disease

End: Akin method

### 3.1.4 Health indicators-categories selection

The given health indicator input is checked to find the categories in the WHO-SEG database, that belongs to the categories can be either of these status, service, risk factors and coverage.

After finding the categories, all the attribute based matched information relevant to categories are identified as crucial collections. Procedure 4 illustrate the details of health indicator category selections.

**Procedure 4: Categories selection**

Auxiliary Variables: Assign PD as Propositions Disease
Assign IQ as Indicators Questions
Assign WHOS as Categories Segment
Assign CA as Categories

Begin:

1. Get PD or IQ from User or Propositions
2. Get all WHOS from WHO-SEG Database
3. If user select IQ (“Life expectancy at birth”) then
4. While read WHOS then
5. If WHOS contains IQ then
6. Select CA, Categories: Risk, Status, Coverage, Services
7. End if
8. End while
9. Else if user select IQ then
10. Go to Traits Section
11. End if

End: Categories Selection

### 3.1.5 Traits based extraction and friendship method

After the categories selection, attributes are selected to extract all information. If the indicator input is “life expectancy at birth” then, their category is “status”. The result should be in the form of numerical or statistical. In the WHO database, the scheme extracts every information regarding the attributes like country, year, and gender (male, female, and both sexes). The friendship method reads all information from the common array, checks associated attributes, and finds the numerical status of each attribute [26]. Finally, the friendship method gives indicator results with respective categories.

**Procedure 5: Traits based extraction and friendship method**

Auxiliary Variables: Assign PD as Propositions Disease
Assign IQ as Indicators Questions
Assign WHOS as Categories Segment
Assign CA as Categories
Assign AT as Attributes
Assign EX as Extraction Details
Assign MC as increment Variable
Assign FC as increment Variable
Assign BC as increment Variable
Assign YC as increment Variable

Begin:

1. If CA is Status Categories, then
2. Read all AT from EMR-Report Database
3. If AT is “Year” or “Gender” or “Country” then
4. While read AT then
5. If Year = ‘1990’ or ‘2000’ or ‘2012’ or ‘2013’ or Gender = ‘Male’ or ‘Female’ or ‘Both’ Then
6. Extract All attributes info and Store EX(n)
7. End If
8. While Read EX then
9. For EX to EX (i1, i2, i3… in)
10. If EX(i) = ‘Male’ then
11. MC = MC + 1
12. Else if EX(i) = ‘Female’ then
13. FC = FC + 1
14. Else if EX(i) = ‘Both’ then
15. BC = BC + 1
16. Else if EX(i) = ‘Year’ then
17. YC = YC + 1
18. If Year = ‘1990’ or ‘2000’ or ‘2012’ or ‘2013’ or Male’ or ‘Female’ or ‘Both’ Then
19. YC = YC + 1
20. End If, while
21. Result Indicators: Display- MC or FC or BC or YC.

End: Traits based extraction and friendship method

Procedure 5 shows the traits based feature extraction using Friendship model. In extend, the proposed work
developed DLAFM with the help of LPHI-CNN model. In this approach, the CAFM is improved with CNN layers and deep training procedures. As given in procedure 6, LPHI-CNN model assists CAFM to improve the accuracy rate by increasing the system learning rate.

Procedure 6: DLAFM- LPHI CNN

Input: Assign auxiliary variables for disease prediction database

Output: Best health indicators derived from the trained values

Begin

1. Initialize the health auxiliary variables (k means clustering method)
2. Get Prediction disease database (raw contents)
3. Prediction database randomly divided into trained data, test data.
4. Consider $\alpha$ is the learning rate and $\beta$ is the hyper parameters values of the prediction Dataset.
   
   $\alpha = \frac{W}{I}$, w-weights updated, I-Iterative events.
5. Data significance rate processing from Experienced Data
   
   $\exists x (\text{Person}(x.1) \land \forall y(\text{Time}(y.1) \rightarrow \text{auxiliaryvalue}(x.1, y.1)))$

6. Inference Rate is defined as
   
   $IR = \frac{dw}{tw}$, dw- number of differential weights produced, tw- total weights produced
7. Fixed the threshold values for the prediction level should be less then $\forall \forall$.
8. Set the CNN layers filters for predicting the health indicators.
9. Incorporate the akin method values in the mathematical model,
   
   If $SR > 5$ and $DIR > 5$ and $LR > 5$
   
   then,

   Result: go to Akin and Traits Based Extraction and Friendship method for best health hypothesis.
10. Repeat the steps until the auxiliary values reaches the threshold value.

End: DLFAM

Procedure 6 illustrates the Learning Rate (LR), Significant Rate (SR), and Deep Inference Rate (DIR) for tuning the CNN model to predict and detect the health indicators. CNN is not only useful for image data analysis but it is a suitable network for analyzing evidence-based hypothetical data analysis cases. Likewise, CNN can be trained with minimal latency to attain optimal decision accuracy in medical diagnosis models. In this DL approach, the increasing learning rate improves the classification accuracy of the proposed HIDSE system. Thus the proposed system improves the overall performance of any medical diagnosis system.

4 Empirical findings and system analysis

This experiment develops various proposed hypothesis models to detect the disease symptoms by extracting health indicators. In this regard, this experiment uses 43,000 ICD-10 records, 1,00,000 patient EMR records, and 20,000 WHO records for the implementation process. The illustration of the proposed method performs hypothesis tests and extracts health indicators. The experiment setup contains different datasets with various health indicators. The health indicator measurements shall be useful to train any computerized medical diagnosis models once the data are effectively tested by optimal hypothetical systems.

The measurements are calculated by the following formulas.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)
\]

TP: True Positive, TN: True Negative, FP: False Positive and FN: False Negative.

The time complexity, measured in seconds taken to complete the hypothesis test. Similarly, the Root Mean Square Error RMSE (RMSE) is defined as the ration between the standard deviation of overall predicted hypothesis errors and the total hypothesis observations. Data reduction rate is important for minimizing the total productions in to lower quantity domain. According to that, the data reduction rate is defined as given in Eq. (5).

\[
\text{Data reduction rate} = \left(\frac{\text{TS} - \text{RS}}{\text{TS}}\right) \quad (5)
\]

TS-Total samples, RS-Removed Samples.

In this regard, this experiment conducts various tests as given in Table 1.

Table 1 illustrates results and time for 10 hypothesis tests, overall the proposed system has taken the test
average time is 0.115 Sec even test is positive or negative. Figure 3 is a graphical representation of Table 1. Figure 3 illustrates the details of hypothesis tests from various trials for identifying the health indicator and disease symptoms [27, 28]. In this case, each hypothesis test takes time in seconds to observe the disease symptoms. Among these observations, three hypothesis results show negative disease symptoms and others show positive symptoms.

In a comparative analysis, the proposed data analysis models are experimented with using different metrics. Figure 4 shows the system classification accuracy of different systems. In this comparison, the proposed DLAFM and CAFM techniques are evaluated against the existing systems like Compositional Medical Health Indicator Analysis (CMHIA), Covid Health Data Analysis (COHDA), and Mental Health Risk Behaviour Analysis (MHRBA). Among these existing systems, CMHIA provides the data analysis model for enabling compositional practices that analyze children’s health indicators. In this approach, the authors developed cross-sectional data analysis procedures to detect the correlations between various data features. In the same manner, the MHRBA system analyses children’s mental health indicators and neural behaviors.

Similarly, CODHA is a recent work that extracts the data features and health indicators of covid patients for developing a medical decision-making system. This existing model helps to understand the dispersion of covid cases. Generally, these models are evolving to build efficient medical assisting platforms. Figure 4 illustrates time varying accuracy rates for different classification systems. The existing systems produce a classification accuracy rate between 93% and 95.5% at different time intervals. At the same time, the proposed CAFM gives a 97% of disease detection accuracy rate, still, it is lower than the proposed DLAFM. In this comparison, the proposed DLAFM is deeply trained with effective data samples of taken datasets. Thus the system works effectively.

Figure 5 provides a detailed comparison between proposed and existing systems with time complexity rates. The time complexity of the systems varies from 0.1 to 1.

In this case, all systems are producing all most equal complexities in time scale. This indicates the algorithms used for analyzing the medical data take equal time overhead to be completed. In contrast, the proposed techniques are dominating other systems’ performances in accuracy and precision. In addition, the proposed DLAFM enables CNN-based medical data hypothesis models for getting more accurate results. In this case, there are few other works identified to get neural observation in dental health indications [29][30][31].

Figures 6 and 7 show the results obtained for system precision and Root Mean Square Error (RMSE). Both performance metrics are showing the system’s ability in detecting the actual diseases correctly and the error produced at run time. This evaluation reveals that the proposed DLAFM and CFAM are having better precisions and minimal RMSE rates than existing techniques. This evaluation procedure has 10 different hypothetical tests on huge medical datasets. In this manner, the RMSE is observed over a varying number of epochs. The production of minimal error rates and better classification accuracy depends upon the handling of raw medical datasets. This proposed system effectively analyses the WHO datasets.
EMR datasets, and ICD-10 datasets. Initially, the raw dataset contains more irrelevant data points, missed health indicators, noises, and repeated medical data. This reduces the quality of any medical hypothesis models. In this regard, this proposed system reduces these portions of raw datasets and initiates optimal data sampling processes.
Thus it controls the performance quality of proposed DLAFM and CAFM techniques. Figure 8 shows the data reduction rates in each dataset.

This shows that the increasing number of epochs increases data reduction quality. Figure 8 indicates the proposed system reduces nearly 30–35% of useless medical data from different datasets.

5 Conclusion

Patient health records and health indicator extraction are the major medical tasks in these computerized diagnosis platforms. In this regard, the proposed HIDSE was a method of medical hypothesis and extracting health indicators it provided clinicians with patients to healthcare sectors. The first part of the proposed method had hypothesis setting, trait selection, evidence check, and hypothesis modeling based on an Akin method for finding a correct hypothesis from a given input. The second part of the proposed method used the Friendship approach that executed input processing, category selection, traits extraction, and health indicators extraction to find the statistical information reports. These methods were called HIDSE, which can be used to improve the patient care system, identify the strengths and weaknesses of therapists, and provide a more consistent database for the understanding of physical therapy practice. In the future, this proposed work will be extended to support multilevel dynamic user queries on real patient data to get results in all aspects with an analytics report.

Acknowledgements This work has been supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia., under grant number: RSP-2021/150.
Author contributions LSK and SR are associated in model formulation and mathematical validations. AVP and SR are associated in paper drafting and writing. VP, AM, and MSA have contributed in data acquisition, simulations and comparisons.

Funding This work has been supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia., under grant number: RSP-2021/150.

Data availability Research data are not shared.

Declarations

Conflict of interest The submitted work is unpublished and the authors confirm no conflict of interest associated with this.

References

1. ICD10: http://www.cdc.gov/nchs/icd/icd10cm.htm. Accessed 06 Apr 2014.
2. EMR boat records. http://www.emrbots.org/. Accessed 29 Aug 2015.
3. WHO death report. http://www.who.int/healthinfo/statistics/mortality_rawdata/en/. Accessed 12 Apr 2014.
4. Shen, F., Liu, S., Fu, S., Wang, Y., Henry, S., Uzuner, O., & Liu, H. (2021). Family history extraction from synthetic clinical narratives using natural language processing: overview and evaluation of a challenge data set and solutions for the 2019 National NLP Clinical Challenges (n2c2)/Open Health Natural Language Processing (OHNLP) competition. JMIR Med Inform. 9(1), e24008.
5. Kim, Y., Heider, P.M., Lally, I.R., Meystre, S.M.: A hybrid model for family history information identification and relation extraction: development and evaluation of an end-to-end information extraction system. JMIR Med Inform. 9(4), e22797 (2021). https://doi.org/10.2196/22797. PMID: 33885370; PMCID: PMC8103307
6. Clotworthy, A., Dissing, A.S., Nguyen, T.-L., Jensen, A.K., Andersen, T.O., Bilsteen, J.F., Elsenburg, L.K., et al.: ‘Standing together–at a distance’: documenting changes in mental-health indicators in Denmark during the COVID-19 pandemic. Scand J Public Health 49(1), 79–87 (2021)
7. Pengpid, S., Peltzer, K.: Associations of number of victimizations with mental health indicators and health-risk behaviours among a nationally representative sample of in-school adolescents in Curacao. Child Abuse Negl. 111, 104831 (2021)
8. Vainselboim, B.: RETRACTED: facemasks in the COVID-19 era: a health hypothesis. Med. Hypotheses (2021). https://doi.org/10.1016/j.mehy.2020.110411
9. Elhadjamor, E.A., Ghannouchi, S.A.: Analyze in depth health care business process and key performance indicators using process mining. Procedia Comput Sci 164, 610–617 (2019). https://doi.org/10.1016/j.procs.2019.12.227
10. Henry, S., Mclnnes, B.T.: Indirect association and ranking hypotheses for literature based discovery. BMC Bioinformatics 20, 425 (2019). https://doi.org/10.1186/s12859-019-2989-9
11. Fan, F.S.: Ractopamine residue in meat might protect people from Parkinson disease. Med. Hypotheses 145, 110397 (2020). https://doi.org/10.1016/j.mehy.2020.110397
12. Zaydman, A.M., Strokova, E.L., Pahomova, N.Y., Gusev, A.F., Mikhailovskiy, M.V., Shevchenko, A.I., Zaidman, M.N., Shilo, A.R., Subbotin, V.M.: Etiopathogenesis of adolescent idiopathic scoliosis: review of the literature and new epigenetic hypothesis on altered neural crest cells migration in early embryogenesis as the key event. Med. Hypotheses 151, 110585 (2021). https://doi.org/10.1016/j.mehy.2021.110585
13. Qin, Q., Jin, B., Liu, Y.: A secure storage and sharing scheme of stroke electronic medical records based on consortium blockchain. BioMed Res Int 2021, 14 (2021)
14. Pai, M.M.M., Ganiga, R., Pai, R.M., et al.: Standard electronic health record (EHR) framework for Indian healthcare system. Health Serv Outcomes Res Method (2021). https://doi.org/10.1007/s10742-020-00238-0
15. Ostheimer, J., Chowdhury, S., Iqbal, S.: An alliance of humans and machines for machine learning: hybrid intelligent systems and their design principles. Technol. Soc. 66, 101647 (2021). https://doi.org/10.1016/j.techsoc.2021.101647
16. Fairclough, S.J., Tyler, R., Dainty, J.R., Dumuid, D., Richardson, C., Shepstone, L., Atkin, A.J.: Cross-sectional associations between 24-hour activity behaviours and mental health indicators in children and adolescents: a compositional data analysis. J Sports Sci 1–13 (2021).
17. Maia, L.O., Daldegan-Bueno, D., Fischer, B.: Opioid use, regulation, and harms in Brazil: a comprehensive narrative overview of available data and indicators. Substance Abuse Treatment Prevention Policy 16(1), 1–10 (2021)
18. Sondik, E., Fielding, J.: Importance of local data and resource allocation for effective successful public health interventions to reduce COVID-19 transmission: commentary on COVID-19 medical vulnerability indicators: a predictive local data model for equity in public health decision-making. Int. J. Environ. Res. Public Health 18(10), 5454 (2021)
19. de Los Rios, P., Okoli, C., Castellanos, E., Allan, B., Young, B., Brough, G., Muchenje, M., et al.: Physical, emotional, and psychosocial challenges associated with daily dosing of HIV medications and their impact on indicators of quality of life: findings from the positive perspectives study. AIDS Behav. 25(3), 961–972 (2021)
20. Skinner, G.C.M., Farrington, D.P.: Self-reported and general practitioner recorded indicators of lifetime health up to age 48 according to offender type in the Cambridge Study in delinquent development. Crim. Behav. Mental Health 31(3), 211–219 (2021)
21. Rosenthal, F.S.: A comparison of health indicators and social determinants of health between Israel and the Occupied Palestinian Territories. Glob. Public Health 16(3), 431–447 (2021)
22. Fischer, B., Daldegan-Bueno, D., Jones, W.: Comparison of crude population-level indicators of opioid use and related harm in New Zealand and Ontario (Canada). Pain Ther. 10(1), 15–23 (2023)
23. Wu, Y., Zhang, C., Liu, H., Duan, C., Li, C., Fan, J., Li, H., et al.: Perinatal depressive and anxiety symptoms of pregnant women during the coronavirus disease 2019 outbreak in China. Am. J. Obstet. Gynecol. 223(2), 240-e1 (2020)
24. Yin, R., Feng, W., Wang, T., Chen, G., Wu, T., Chen, D., Lv, T., Xiang, D.: Concomitant neurological symptoms observed in a patient diagnosed with coronavirus disease 2019 in China. Am. J. Obsetr. Gynecol. 223(2), 240-e1 (2020)
25. Nobel, Y.R., Hippius, M., Zucker, J., Lebwohl, B., Wang, T.C., Fuentes, D., et al.: A systematic review and meta-analysis. Clin. Res. Hepatol. Gastroenterol. 44(5), 653–661 (2020)
26. Wang, H., Qiu, P., Liu, J., Wang, F., Zhao, Q.: The liver injury chain. BioMed Res Int 2021, 14 (2021)
27. Rahmani, A.M., Babaei, Z., Souri, A.: Event-driven IoT architecture for data analysis of reliable healthcare application using complex event processing. Clust. Comput. 24(2), 1347–1360 (2021)
28. Kahlessenane, F., Khaldi, A., Kafi, R., Euschi, S.: A robust blind medical image watermarking approach for telemedicine applications. Clust. Comput. 24(3), 2069–2082 (2021)
29. Cheng, J., Liao, Y., Dong, Y., Hu, H., Yang, N., Kong, X., Li, S., et al.: Microglial autophagy defect causes parkinson disease-like symptoms by accelerating inflammasome activation in mice. Autophagy 16(12), 2193–2205 (2020)
30. Peball, M., Krismer, F., Knaus, H.-G., Djamshidian, A., Werkmann, M., Carbone, F., Ellmerer, P., et al.: Non-motor symptoms in Parkinson’s disease reduced by Nabilone. Ann. Neurol. 88(4), 712–722 (2020)
31. Mahmoud, N.M., Fouad, H., Alsadon, O., Soliman, A.M.: Detecting dental problem related brain disease using intelligent bacterial optimized associative deep neural network. Clust. Comput. 23(3), 1647–1657 (2020)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A.V. Prabu is an Associate Professor in Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Andhra Pradesh, India. He obtained his B.E degree from Anna University, Tamil Nadu in 2005. He completed his Master degree from BPUT, Odisha, India. Presently Pursuing Ph.D. in JNTUK, Kakinada, Andhra Pradesh, India. He has 14 years of academic experience in Electronics and Communication Engineering and Electronics and Computer Science Engineering domains. His research interests include wireless sensor networks, wireless communication and internet of things.

L. Sathish Kumar is the Assistant Professor in Gaming Division, School of Computing Science and Engineering at VIT Bhopal University, Bhopal. He has completed his Ph.D., in Computer Science and Engineering, for Alagappa University, Karaikudi. He has 8 years of academic experience and has more than 29 research publications in refereed International/National Journals and Conferences. He has 2 international patents and 1 design patent to his credit and has received 2 awards for his research and academic contributions. He is an active researcher in the field of Medical Image Processing, Deep Learning, Machine Learning, and Data Science.

S. Rajasoundaran is an Assistant Professor in the School of Computer Science and Engineering, VIT University, Bhopal, Madhya Pradesh, India. He obtained his Ph.D in information Science and Communication Engineering from Anna University, Chennai, India. He completed his Master degree in Network Engineering. He has 12 years of academic experience in computer science and information technology domains. He has published numerous reputed journals in his fields. His research interests include computer networks, image processing, wireless networks and network security.

S. Pandimurugan Currently working as Assistant professor in the department of networking and communications, SRMIST, Chennai. He has completed his Ph.D. in Computer Science & Engineering in 2019 from Manonmaniam Sundaranar University (MSU) Tirunelveli. He has done his B.E, in EEE and M.E in CSE from Anna University in the year 2005 and 2008 respectively. Has 11 years of teaching experience and has 22 publications in International Journals and Conferences. His research interests include Health Monitoring, Cyber Physical Systems and IoT. He is an active member of ACM and has authored the book titled-Principles of Communication, in the year 2012, published by Bharathi Publishers Chennai. He published one Scopus indexed book chapter and filed and published two international patents also one Indian patent to his credit.
Amrit Mukherjee is currently working as an Assistant Professor (II) in School of Electronics, KIIT DU Odisha, India. He was a Post-Doc Research Fellow in School of Computer Science and Communication Engineering, Jiangsu University, PR China from May 2018-June 2020. He has worked in C-DAC (Armenia and Ghana) as Project Engineer. He has published more than 70 research articles, 1 patent published and submitted 5 patents filed under review. Also, he has been a reviewer in many IEEE journals and other prestigious conferences. He also served special issue guest editor for Computer Communications: Multimedia Tool and Applications; Journal of System Architecture; Computers, Material and Continua; and IEEE Internet of Things Journal. His area of interests includes Artificial Intelligence, next generation IoT systems, Wireless Sensor Networks, Cognitive Radio, and signal processing.

Mohammed S. Al-Numay was born in Riyadh, Saudi Arabia. He received the B.S. degree (with honors) from King Saud University, Riyadh, Saudi Arabia, in 1986, the M.S. degree from Michigan State University, East Lansing, MI, in 1990, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, GA, in 1997, all in electrical engineering. Since 1998, he has been with the Electrical Engineering Department at King Saud University, where he is now a Full Professor. During the period 2002–2006, he was the Dean of Admissions and Registration at King Saud University. From 2008 until now, he is a Senior Consultant of Student Information Systems (SIS) and electronic admission to many governmental and private universities and colleges. He was appointed as the Assistant Vice Rector for Information Technology from 2015 to 2017. In January 2018 he became King Saud University Vice President for Educational and Academic Affairs. His research interests include modeling and control of switched mode power supplies.

Authors and Affiliations

L. Sathish Kumar1 · Sidheswar Routray2 · A. V. Prabu3 · S. Rajasoundaran1 · V. Pandimurugan4 · Amrit Mukherjee5 · Mohammed S. Al-Numay6

1 School of Computing Science and Engineering, VIT University, Bhopal, India
2 Department of Computer Science and Engineering, School of Engineering, Indrashil University, Rajpur, Mehsana, Gujarat, India
3 Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India
4 School of Computing, Department of Networking and Communications, SRMIST, Kattankulathur Campus, Chennai 603203, India
5 Department of Computer Science, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic
6 Department of Electrical Engineering, King Saud University, Riyadh, Saudi Arabia