PREDICTIVE ANALYTICS MODELS FOR ELECTRONIC HEALTH RECORD (LITERATURE REVIEW)

Yuda Syahidin, Pemi Novita Sari,
Manajemen Informasi Kesehatan, Piksi Ganesha Politechnic, Bandung, Indonesia

Abstract

Big data growth in the healthcare community, accurate analysis of medical data supports early disease detection, patient care and community services. However, the accuracy of the analysis decreases when the quality of the medical data is incomplete. In addition, different regions show unique regional disease characteristics, which can weaken the prediction of disease outbreaks. EHR is designed to store patient medical information. Predictive Analytics involves a variety of techniques from modeling, machine learning, and data mining that break down past and present data features to predict the future of medical record data. In this survey paper, we discuss predictive analytical models in the field of EHR that have been made by previous researchers. The purpose of this paper is to provide an overview of future research opportunities in building a Predictive Analytics Model for Electronic Health Records

Keywords: Predictive Analytics, Electronic Health Record, Machine Learning

Introduction

The rapid development of the internet, electronic medical information has become popular in all cities and around the world, such as electronic health records (EHR) to replace paper medical records, online appointments, and online reports, so that the volume of data growth is growing and many have encouraged research. intensive in the development of technological devices to store, manage, and analyze data in the health sector (Hong dkk, 2019). Even though there is still a large portion of health data that is still not utilized, for this it is necessary to carry out Unstructured Data Analysis to become the next innovation in data science in the health sector. Scalability and efficiency in analyzing medical record data are still lacking in generating Electronic Health Reports, there is still a lack of a decision-making system in the health sector because there is still unstructured health data.

The use of Electronic Health Records (EHR) makes it possible to analyze large amounts of medical data (Shickel dkk, 2017). Recently Deep Learning Techniques can play an important role in managing the huge medical data that has been generated every day (Ismail dkk, 2020; Xu dkk, 2020)]. In addition, deep learning achieves success in multiple fields by effectively constructing deep hierarchical features.

Predictive analytics is one of the critical areas of clinical science to offer better care to patients. Predictive Analytics helps optimize resource costs and provide better care to patients in resource-constrained environments. This helps make the diagnosis of doctors in providing patients with faster and better treatment. methods of translating current data into new, unexpected situations and environments that can contribute to a more in-depth knowledge of disease-related information, such as better monitoring of different phases of disease, and identifying disease onset. The result is improved quality of care to patients, increased support for physicians, and easy validation of standard diagnostic procedures for disease, and provision of healthcare services to patients. In recent years, most of the methods used to evaluate large EHR data for predictive models are based on machine learning and standard statistical techniques used in making informed decisions about complex issues such as clinical trial results (Jensen dkk, 2012), as well as in predictive analysis to determine the likelihood of predicting patient re-admission to the hospital.
The Electronic Health Record (EHR) is an increasingly common data source for clinical risk prediction, presenting unique analytical opportunities and challenges [3]. The thing that needs to be considered and is a challenge in making predictions about Electronic Medical Record data is that predictive algorithms have good performance depending on data representation and selection of appropriate features. Another challenge in feature selection is analyzing, selecting, and evaluating raw EHR data which can be time-consuming and often requires trials and determination of possible predictive variables in EHR potentially in the thousands, especially when clinical records are unclear from doctors,

nurses, and service providers. others used (Goldstein dkk, 2017; Weiskopf dkk, 2013). Based on this, it is necessary to develop a predictive model that will later perform pattern recognition, statistics, databases, and visualization for handling the problem of retrieving information from a larger health database.
In this paper, we will discuss the predictive analytics model by previous research, as well as the model that will be examined based on previous research opportunities.

Electronic Health Records

An Electronic Health Record (EHR) is a person's digitally stored health information and is instantly and securely available to authorized users. EHRs contain patient

diagnoses, medications, vital signs, treatment plans, progress notes, radiological images, and test results [2]. Classification of records exists as unstructured and structured EHR data. Unstructured EHR data is written or dictated based on the clinical context that describes the patient's condition and is most useful for clinical documentation. However, because of their unstructured format, numerous typos and spelling errors, and use of acronyms, abbreviations, and oddities, they pose a challenge to computer analysis. Classification of structured RKE data can be as administrative data and additional data. Administrative data either remain unchanged throughout the clinical encounter (such as demographic data) or continue to change over time (such as diagnoses and procedures). Additional data may be discrete (such as physiological measurements, medications, and laboratory tests) or continuous (such as respiration and blood pressure).

EHR data consists of various types of data, from structured information such as drug prescription data consisting of dates and doses captured through standard prescribing systems, to unstructured data such as clinical narratives that describe the medical reasons behind medical record documents. The relationship between structured data and unstructured data can be seen in Figure 1 below.

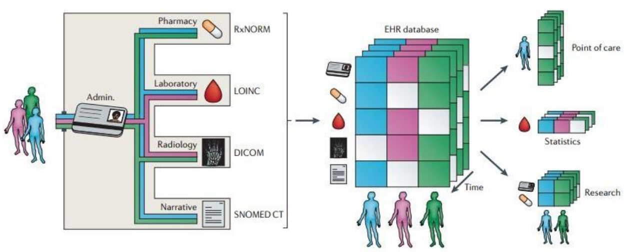


Figure 1. Electronic health record content (Jensen dkk, 2012)

A patient's Electronic Health Record (EHR) can be viewed as a storehouse of information regarding his health status in a computer-readable form. When connected to the health care system it produces various types of patient-related data. Patient data is stored in a database and can be viewed in a format that suits the needs and authority of certain user groups. For example, a doctor may request EHR data for a particular patient, statistical summaries of all laboratory procedures and data extraction used for research.

The term Electronic Health Record (EHR), or Electronic Medical Record (EMR), refers to the collection of patient health information in a digital format. EHRs can be categorized in terms of functionality: (i) basic EHR without clinical records, (ii) basic EHR with clinical records and (iii) comprehensive systems. EHRs, even in their simplest form, provide researchers with a rich collection of data. Data can be shared across networks and can include, as previously described, a wide variety of information. EHR is primarily designed for internal hospital administration tasks and many different schemes exist in different structures(Liang dkk, 2019).

Results and Discussion

Predictive Analytics consists of various techniques that use current data and historical data to predict future results.

Predictive Analytics seeks to find relationships in data and uncover patterns.

To perform Predictive Analytics requires a formula or set of rules applied to derive a score or code, which is known as an analytical model. This analytic modeling uses mathematical techniques to identify meaningful relationships between variables (Leventhal, 2018). Each model used is a simplification of the existing reality, which can help in understanding problems and making predictions.

There are several models used for Predictive Analytics, including:

A. Decision Tree Analysis.

Decision trees are good predictive models and have many advantages, including being easy to understand and can be presented in very simple or detailed ways. The Decision Tree will start at the root of the tree, which represents all the data, and continue to break each category into two separate categories based on the optimal method, to be able to guess the best characteristics to identify each of the two categories (splitting a node). Then the algorithm will continue to divide the data until it can no longer divide, or stop in another way based on the control parameters (Leventhal, 2018).

B. Random Forest

Random Forest is an extension of the decision tree approach, which was developed by Breiman (Futoma dkk, 2015).

Random Forest is one of the methods in the Decision Tree. A decision tree or decision-making tree is a flowchart that is shaped like a tree which has a root node that is used to collect data, an inner node that is at the root node that contains questions about data and a leaf node that is used to solve problems and make decisions. decision. The decision tree classifies a data sample whose class is unknown into existing classes. The use of a decision tree to avoid overfitting a data set while achieving maximum accuracy [20].

C. Regression Analysis

Regression analysis is a statistical process used to analyze the relationship between variables, especially to predict one or more target variables (Hao dkk, 2019).

D. Supervised Neural Network

Supervised Neural Networks are a good solution when the relationship between the target and predictor variables is complex and unknown (Sidey-Gibbons & Sidey-Gibbons, 2019).

Neural Network processing basically mimics biological neurons, as seen in Figure 2 (M. Li dkk, 2021).

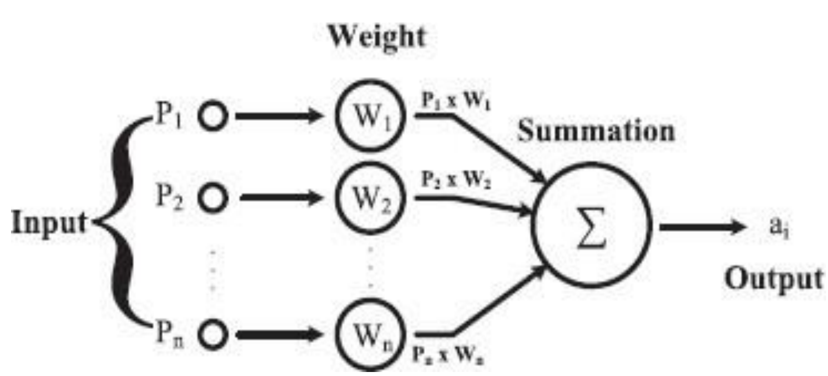


Figure 2. Artificial Neural Network Satu Layer (Sharma & Shah, 2021)

A. Result and Discussion Machine Learning

The Machine Learning method is most commonly used to build medical database systems from EHRs for patients who have undergone health checks (Hauskrecht dkk, 2013), Machine learning methods are still used by previous studies in making predictions that tend to use one or a combination of several algorithms (Sohn Dr. dkk, 2013). The following is previous research using machine learning techniques.

1) S. Mohan [14]

Machine Learning has proven effective in assisting in making decisions and predictions from the large amount of data generated by the healthcare industry. In 2019, S. Mohan (Mohan dkk, 2019) proposes a new method that aims to find significant features by applying machine learning techniques that result in increased accuracy in cardiovascular disease prediction. Predictive models are introduced with a combination of different features and some known classification techniques. prediction model for heart disease with the linear Hybrid Random Forest Model (HRFLM). like picture 2 below.

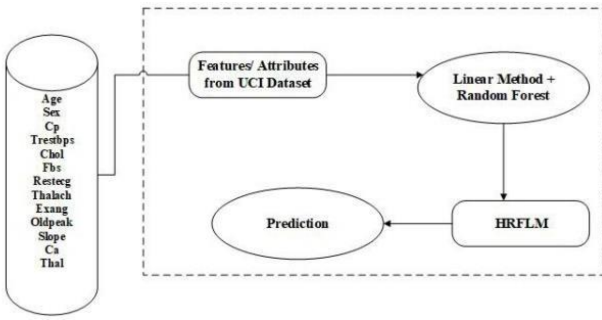


Figure 3. HRFLM Model (Mohan dkk, 2019)

2) K. Yu and X. Xie (Wang dkk, 2021)

This research (Meng dkk, 2020) identifies patients who are at high risk and enable doctors to take action early. This technique uses predictive analytics based on electronic health records (EHR) for readmission of patients to the hospital. Based on this, a hospital readmission framework is proposed. This approach uses inpatient administration data. hospitals from the national health care dataset. The Hospital Readmission framework model is as shown in Figure 3 below.

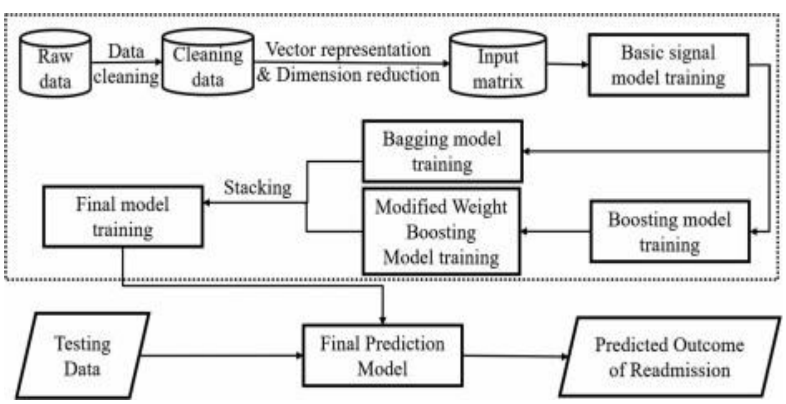


Figure 4. Model Framework Hospital Readmision (Nguyen dkk, 2017)

3) Y. Sun et al (Sun & Zhang, 2019)

Research by Y.Sun (Sun & Zhang, 2019) presents a set of Machine Learning models to diagnose Diabetic Retinopathy (DR) in patients in EHR data and form a set of treatment methods. The method used with the 5 Algorithm techniques in the Machine Learning model, as shown in Figure 4 below.

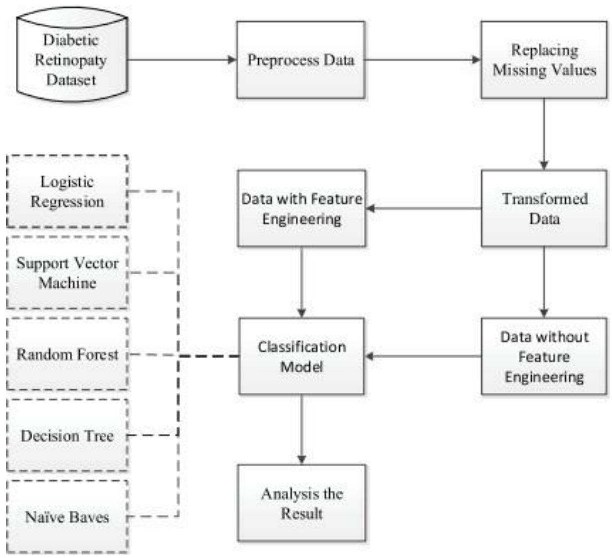


Figure 5. Model Framework Diabetic Retinopathy (Sun & Zhang, 2019)

4) R. Ghorbani (Ghorbani dkk, 2020)

The new hybrid model uses a Genetic Algorithm as a feature selection method and an ensemble model based on a combination of Stacking and Boosting.

This process selects an optimal subset of the relevant features for use in predictive model development. Especially, feature selection techniques can reduce the dimensionality of the dataset by ignoring features that are insignificant or have noise so that predictive models can be more accurate.

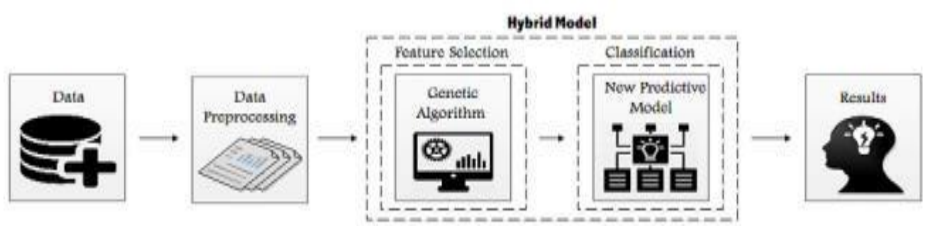


Figure 6. Model Framework New Hybrid (Ghorbani dkk, 2020)

B. Deep Learning

The use of Deep Larning to analyze EHRs data has increased over the last few years; continued growth is facilitated by the availability of more EHRs data, Developments in Deep Learning, and innovative ways to combine these two trends. The ultimate Deep Learning architecture for studying efficient patient representation to predict admission to the emergency room and heart failure patients (Guller & Guller, 2015). A Neural Network (NN) can be represented by a directed graph form which the input layer takes in a signal vector and one or more hidden layers process the output of the previous layer. Deep learning models are an approach that aims to produce end-to-end systems that learn from raw data and perform specialized tasks without supervision. Deep Neural Network presents more layers and more nodes in each layer with respect to neural network. The number of parameters to be set to increase the performance of neural networks and, in addition, deep learning cannot be learned without enough data and without a powerful computer.

There are different types of deep neural architectures such as: (i) Convolutional Neural Networks (CNN) where the pattern of connectivity between layers is inspired by the organization of the visual cortex; (ii) Recurrent Neural Network (RNN) in which the connections between nodes form a directed graph along the sequence and [34] (iii) Feedforward Neural Network (FFN) where information moves in only one direction, forward, from input nodes, through hidden nodes (if any) and finally to output nodes. There are no cycles or loops in the network. Deep learning algorithms are built on top of an Artificial Neural Network (ANN) in which a few interconnected nodes (neurons) are arranged in several different layers. Hidden nodes are layers that do not belong to the input or output layers (Wanyan dkk, 2020).

1) Miotto, R et al (Miotto dkk, 2016)

This model (Miotto dkk, 2016) proposes the Deep Patient model: an unsupervised Deep Learning approach to extract hierarchical features and patterns from EHRs data. This research shows that unsupervised Deep Learning can be used to obtain hierarchical features based on EHR datasets.

A model that represents the predictive results of patient status from EHRs data using deep learning techniques. The purpose of this representation is to predict the patient's health status by determining the likelihood that the patient is currently developing other diseases. They use autoencoder techniques to find a form of hierarchical representation in the dataset in Figure 5 below.

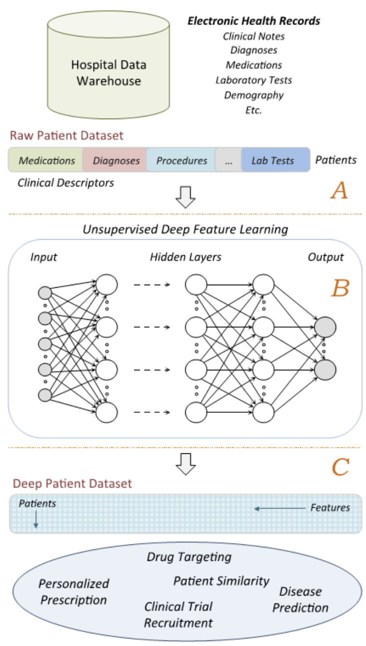


Figure 7. Model Deep Patient (Miotto dkk, 2016)

2) M. Jamshidi et al. (Jamshidi dkk, 2020)

The first step is data compilation. The data being discussed here consists of medical information, such as clinical reports, notes, images and other forms of information that can be converted into machine- understandable data. Human Intervention as part of the machine learning method, takes place and investigates and analyzes data to extract data with structure, patterns, and features. Like figure 8 below.

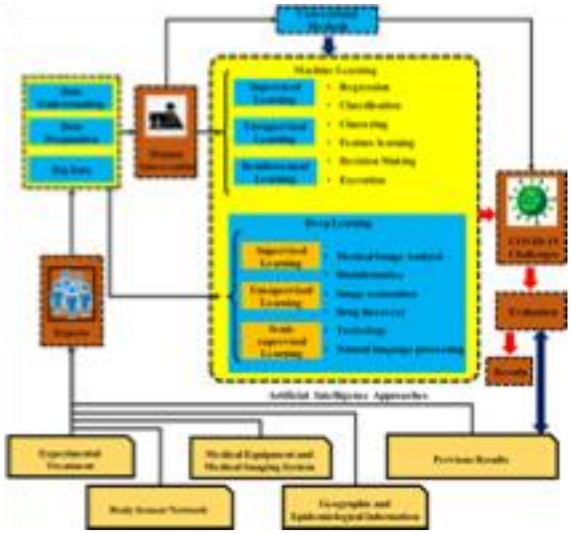
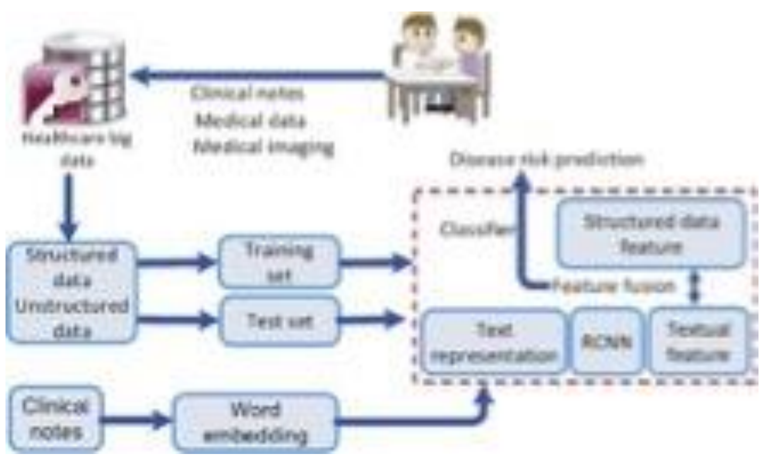


Figure 8. Deep Learning for Diagnosis and Treatment (L. F. Li dkk, 2020)

3) Yixue Hao (Hao dkk, 2019)

In the research conducted, he proposed the MD-RCNN model using multi-modal data to predict disease risk. The multimodal data-based recurrent convolutional neural network (MD-RCNN) model works for disease risk prediction by extracting structured and unstructured features to obtain a highly non-linear relationship between structured and unstructured data. As shown in Figure 9 below.



Gambar 9. Model MD-RCNN (Hao dkk, 2019)

4) J. Baek et al (Çelik dkk, 2018)

The model works based on selected context variables, high-relationship context information is extracted, and then used as DNN-context input for learning. After the data collected is processed, the data is divided into a training set and a test set. For data feature extraction, the data is entered in the DNN, then the data features are extracted and displayed in the hidden layer. The output data is classified with the aim of evaluating the goodness of the model, this model is compared to other models, as shown in Figure 10 below.

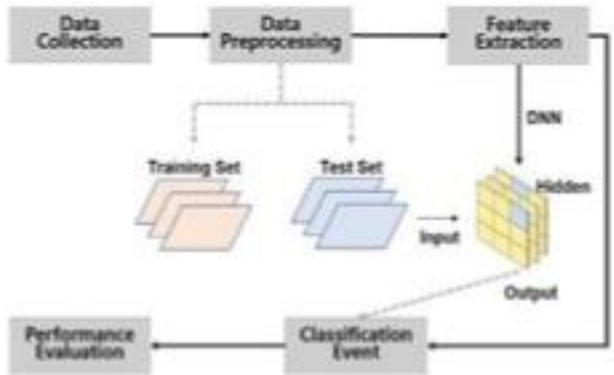


Figure 10. Model Context DNN (Baek & Chung, 2020)

Conclusion

Based on the review paper that has been described in the Predictive Analytics model from previous research, it can be seen the methods used. An explanation of previous research related to Predictive Analytics in the field of electronic health records can be seen in Table 1 below.

Table I.
Comparison Of Model Predictive Analytics In Electronic Health Records

Paper	Method	Focus	Performance
S. Mohan (Mohan dkk, 2019)	Hybrid Random Forest with model linier (HRFLM).	Heart Disease	Make predictions of heart disease up to an accuracy level of 88.7%
K. Yu and X. Xie (Yu & Xie, 2020)	The joint ensemble- learning model, Modified Weight Boosting- Stacking (MWBS)	Return of the patient to the hospital	Make predictions about patient arrivals up to an accuracy level of 89.1%
Y. Sun (Sun & Zhang, 2019)	Classification Model: Logistic Regression (LR), Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), Naïve Bayes (NB)	Diabetic retinopathy (DR) – Eye disorder due to diabetes	Predict eye disorders using 5 (five) classification algorithms and produce predictions with RF up to 92%
R. Ghorbani (Ghorbani dkk, 2020)	Genetic Algorithm	Mortality	This model achieves 98.20% by 88.47% with the AUC test
Miotto, R et al (Miotto dkk, 2016)	Unsupervised deep feature learning	78 kinds of disease (disease)	Generate predictive results of more than one kind of prediction
P. Nguyen et al (Sabokrou dkk. 2018	Multilayered architecture based on CNN	Clinical Notes, Medical Codes	Hospital re-admission prediction
Yixue Hao (Hao dkk, 2019)	MD-RCNN algorithm	Clinical notes, Medical data, Medical Imaging	Melakukan prediksi terhadap resiko penyakit berdasarkan data kesehatan yang terstruktur maupun tidak terstruktru
J. Baek et al [36], 2020	Deep Neural Network dan Multiple Regression	Depression	Nilai yang diprediksi adalah nilai antara 0 dan 1. Risiko depresi memiliki empat tahap, yaitu 'baik', 'tidak buruk', 'berisiko', dan 'sangat berisiko'

Berdasarkan penjelasan di penelitian–penelitian sebelumnya bahwa tantangan dalam melakukan Predictive Analytics diantaranya data yang tidak terstruktur dan terstruktur. Oleh karena itu, salah satu tantangan utama adalah mengintegrasikan dan menyelaraskan data yang memiliki perbedaan. Selain itu sifat heterogen dari tipe data termasuk data numerik, objek waktu tanggal, free text (hasil klinis) menimbulkan tantangan signifikan untuk mendapatkan informasi penting dalam EHRs serta dikarenakan data EHRs dalam bentuk struktur dan tidak terstruktur memungkinkan diperlukan model klasifikasi dan klustering untuk menentukan predictive analytics dibidang Rekam Medis Elektronik.

References

Baek, J. W., & Chung, K. (2020). Context Deep Neural Network Model for Predicting Depression Risk Using Multiple Regression. IEEE Access, 8, 18171–18181. https://doi.org/10.1109/ACCESS.2020.2968393

Çelik, A., Yaman, H., Turan, S., Kara, A., Kara, F., Zhu, B., Qu, X., Tao, Y., Zhu, Z., Dhokia, V., Nassehi, A., Newman, S. T., Zheng, L., Neville, A., Gledhill, A., Johnston, D., Zhang, H., Xu, J. J., Wang, G., … Dutta, D. (2018). FUNDAMENTALS OF MACHINE LEARNING FOR PREDICTIVE DATA ANALYTICS. In Journal of Materials Processing Technology (Vol. 1, Nomor 1). http://dx.doi.org/10.1016/j.cirp.2016.06.001%0Ahttp://dx.doi.org/10.1016/j.powtec.2016.12.055%0Ahttps://doi.org/10.1016/j.ijfatigue. 2019.02.006%0Ahttps://doi.org/10.1016/j.matlet.2019.04.024%0Ahttps://doi.org/10.1016/j.matlet.2019.127252%0Ahttp://dx.doi.o

Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. Journal of Biomedical Informatics, 56, 229–238. https://doi.org/10.1016/j.jbi.2015.05.016

Ghorbani, R., Ghousi, R., Makui, A., & Atashi, A. (2020). A New Hybrid Predictive Model to Predict the Early Mortality Risk in Intensive Care Units on a Highly Imbalanced Dataset. IEEE Access, 8, 141066–141079. https://doi.org/10.1109/ACCESS.2020.3013320

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. Journal of the American Medical Informatics Association, 24(1), 198–208. https://doi.org/10.1093/jamia/ocw042

Guller, M., & Guller, M. (2015). Big Data Technology Landscape. Big Data Analytics with Spark, Ml, 1–15. https://doi.org/10.1007/978-1-4842-0964-6_1

Hao, Y., Usama, M., Yang, J., Hossain, M. S., & Ghoneim, A. (2019). Recurrent convolutional neural network based multimodal disease risk prediction. Future Generation Computer Systems, 92, 76–83. https://doi.org/10.1016/j.future.2018.09.031

Hauskrecht, M., Batal, I., Valko, M., Visweswaran, S., Cooper, G. F., & Clermont, G. (2013). Outlier detection for patient monitoring and alerting. Journal of Biomedical Informatics, 46(1), 47–55. https://doi.org/10.1016/j.jbi.2012.08.004

Hong, W., Xiong, Z., Zheng, N., & Weng, Y. (2019). A Medical-History- Based Potential Disease Prediction Algorithm. IEEE Access, 7, 131094–131101. https://doi.org/10.1109/ACCESS.2019.2940644

Ismail, W. N., Hassan, M. M., Alsalamah, H. A., & Fortino, G. (2020). CNN-based health model for regular health factors analysis in internet-of-medical things environment. IEEE Access, 8, 52541– 52549. https://doi.org/10.1109/ACCESS.2020.2980938

Jamshidi, M., Lalbakhsh, A., Talla, J., Peroutka, Z., Hadjilooei, F., Lalbakhsh, P., Jamshidi, M., Spada, L. La, Mirmozafari, M., Dehghani, M., Sabet, A., Roshani, S., Roshani, S., Bayat-Makou, N., Mohamadzade, B., Malek, Z., Jamshidi, A., Kiani, S., Hashemi- Dezaki, H., & Mohyuddin, W. (2020). Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment. IEEE Access, 8(December 2019), 109581–109595. https://doi.org/10.1109/ACCESS.2020.3001973

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics, 13(6), 395–405. https://doi.org/10.1038/nrg3208

Leventhal, B. (2018). Predictive Analytics for marketers. Kogan Page.

Li, L. F., Wang, X., Hu, W. J., Xiong, N. N., Du, Y. X., & Li, B. S. (2020). Deep Learning in Skin Disease Image Recognition: A Review. IEEE Access, 8, 208264–208280. https://doi.org/10.1109/ACCESS.2020.3037258

Li, M., Liu, Y., Liu, X., Sun, Q., You, X., Yang, H., Luan, Z., Gan, L., Yang, G., & Qian, D. (2021). The Deep Learning Compiler: A Comprehensive Survey. IEEE Transactions on Parallel and Distributed Systems, 32(3), 708–727. https://doi.org/10.1109/TPDS.2020.3030548

Liang, H., Tsui, B. Y., Ni, H., Valentim, C. C. S., Baxter, S. L., Liu, G., Cai, W., Kermany, D. S., Sun, X., Chen, J., He, L., Zhu, J., Tian, P.,Shao, H., Zheng, L., Hou, R., Hewett, S., Li, G., Liang, P., … Xia, H. (2019). Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nature Medicine, 25(3), 433–438. https://doi.org/10.1038/s41591-018-0335-9

Meng, L., Dong, D., Li, L., Niu, M., Bai, Y., Wang, M., Qiu, X., Zha, Y., & Tian, J. (2020). A Deep Learning Prognosis Model Help Alert for COVID-19 Patients at High-Risk of Death: A Multi-Center Study. IEEE Journal of Biomedical and Health Informatics, 24(12), 3576– 3584. https://doi.org/10.1109/JBHI.2020.3034296

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports, 6. https://doi.org/10.1038/srep26094

Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542–81554. https://doi.org/10.1109/ACCESS.2019.2923707

Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. (2017). Deepr: A Convolutional Net for Medical Records. IEEE Journal of Biomedical and Health Informatics, 21(1), 22–30. https://doi.org/10.1109/JBHI.2016.2633963

Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., & Klette, R. (2018). Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Computer Vision and Image Understanding, 172(January), 88–97. https://doi.org/10.1016/j.cviu.2018.02.006

Sharma, T., & Shah, M. (2021). A comprehensive review of machine learning techniques on diabetes detection. Visual Computing for Industry, Biomedicine, and Art, 4(1). https://doi.org/10.1186/s42492-021-00097-7

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. arXiv, 22(5), 1589–1604.

Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. BMC Medical Research Methodology, 19(1), 1–18. https://doi.org/10.1186/s12874-019-0681-4

Sohn Dr., S., Wagholikar, K. B., Li, D., Jonnalagadda, S. R., Tao, C., Elayavilli, R. K., & Liu, H. (2013). Comprehensive temporal information detection from clinical text: Medical events, time, and TLINK identification. Journal of the American Medical Informatics Association, 20(5), 836–842. https://doi.org/10.1136/amiajnl-2013-001622

Sun, Y., & Zhang, D. (2019). Diagnosis and Analysis of Diabetic Retinopathy Based on Electronic Health Records. IEEE Access, 7, 86115–86120. https://doi.org/10.1109/ACCESS.2019.2918625

Wang, Q., Wang, S., Wei, B., Chen, W., & Zhang, Y. (2021). Weighted K- NN Classification Method of Bearings Fault Diagnosis with Multi- Dimensional Sensitive Features. IEEE Access, 9, 45428–45440. https://doi.org/10.1109/ACCESS.2021.3066489

Wanyan, T., Vaid, A., De Freitas, J. K., Somani, S., Miotto, R., Nadkarni, G. N., Azad, A., Ding, Y., & Glicksberg, B. S. (2020). Relational Learning Improves Prediction of Mortality in COVID-19 in the Intensive Care Unit. IEEE Transactions on Big Data, December 2020. https://doi.org/10.1109/TBDATA.2020.3048644

Weiskopf, N. G., Hripcsak, G., Swaminathan, S., & Weng, C. (2013). Defining and measuring completeness of electronic health records for secondary use. Journal of Biomedical Informatics, 46(5), 830– 836. https://doi.org/10.1016/j.jbi.2013.06.010

Xu, D., Hu, P. J. H., Huang, T. S., Fang, X., & Hsu, C. C. (2020). A deep learning–based, unsupervised method to impute missing values in electronic health records for improved patient management. Journal of Biomedical Informatics, 111(October). https://doi.org/10.1016/j.jbi.2020.103576

Yu, K., & Xie, X. (2020). Predicting Hospital Readmission: A Joint Ensemble-Learning Model. IEEE Journal of Biomedical and Health Informatics, 24(2), 447–456. https://doi.org/10.1109/JBHI.2019.2938995