DOI: 10.15849/IJASCA.230720.13

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 2, July 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523
Copyright © Al-Zaytoonah University of Jordan (ZUJ)

Enhancement the Susceptibility of Developing Diabetes Treatment by Time Series Data Analysis

Leenah Aljuhani[1], Alaa Sagheer[2]

[1] College of Computer Science and Information Technology
King Faisal University, Saudi Arabia
e-mail: 221445351@student.kfu.edu.sa

[2] College of Computer Science and Information Technology
King Faisal University, Saudi Arabia
e-mail: sagheer@kfu.edu.sa

Abstract

Diabetes: This term includes a number of disorders that are characterized by problems with the hormone insulin, which the pancreas naturally produces to help the body absorb the sugar in the blood, as issues in the production of this hormone raise the level of sugar in the blood and may lead to many other diseases and problems. That may lead to death. We use technology in the field of artificial intelligence to help medical specialists to classify diseases of humans by taking some measurements such as blood pressure, diabetes, heart rate, and other readings according to the diagnosis of the disease within periods of time and working on analyzing these readings to visualize the patient's health state. And may analysis to determine the patient's ability to respond to the system. In this paper, we used a Wireless sensor network technology to transmit patient data to the training server, then used artificial intelligence (AI) in the training and classification phase to find out if a patient has diabetes or not.

Keywords: *Diabetes Treatment, Diabetes Mellitus, Classification, Machine Learning, Time Series Data Analysis, DM, TB, AI, technology, screening, machine Learning, WSN.*

1 Introduction

Most of the used applications today such as medical, industrial, business information, or even military applications need a level of AI to be the most accurate working. [1] In order to use AI in the medical field, we will use machine learning to classify the diseases of humans by taking some measurements such as blood pressure, diabetes, heart rate, and other readings according to the diagnosis of the disease within periods of time, then transmission these data using Wireless Sensor Network (WSN) to a base station and working on analyzing these readings to visualize the patient's health state. And may analysis to determine the patient's ability to respond to the system. Where we expect, through our study, that the system will give the results better than the manual results that doctors are doing at the present time. [2]. the analysis of DM time series data will be based on the idea that the accumulation of the effects of lifestyle events such as ingestion and exercise can affect personal health conditions with some delay. During the analysis, the accumulation of the effects of lifestyle events should be represented by a summation of energy supply or expenditure data (kilojoules) due to ingestion or exercise. The

Received 4 March 2023; Accepted 29 May 2023

accumulation of these effects may cause variation of health data such as BMI, BGLs, BP and body-fat percentage with some delay.

Remote patient monitoring (RPM) is a one type of homecare that allows patients to use mobile medical devices and technology to gather patient statistical data and send these data to a cloud server for healthcare professionals [3] . Data that can be collected include vital signs, weight, blood pressure and heart rate. Once collected, patient data is sent to a doctor office through wireless media such as Wireless sensor network (WSN) and a software that can be installed on local devices [4].

The main idea of this thesis is to apply artificial intelligence algorithms to neural networks and other algorithms in the training and classification stage and compare these algorithms to find the best and fastest results for detecting Diabetes and the ability to respond to the drugs [5].

The system makes periodic readings for the patient for some measurements such as pressure, sugar, and other readings necessary for diabetes, then the wireless sensor will take these data and transmit them using wireless nodes through wireless media until arriving at the base station, the base station is connected to the server database that stored the patient data it can allow the patient to view his diagnostic. [6]. then the system sends these readings to the training system, which in turn trains the machine on these readings and finds the difference in them periodically, and then calculates the blood sugar percentage and how different it is from The normal rate and in the event of finding a difference or a significant increase from the normal percentage, the system sends an alarm asking for help by giving a dose of medicines according to the patient's status, and then taking new measurements to know the patient's ability to respond to these medicines. The system learns through the neural network as well as the SVM to compare the speed of extraction of results and the accuracy of extraction, [7] the algorithm has also been compared with other algorithms such as XGboost Algorithm.

2 Related Work

The proposed research's domain has been the subject of numerous research articles published in the literature. For instance, using the available data from 2006 to 2016, the authors in **[8]** used the Box-Jenkins approach to determine the number of diabetes patients for the upcoming two years. The statistical axioms and fundamentals form the foundation of the Box-Jenkins forecasting models. However, the authors did not immediately update the measurements in their work to reflect changes in patient cases, so we cannot find real-time patient results.

The authors in **[9]** introduced attempts to predict the risk of heart disease more accurately using ensemble learning techniques. Moreover, the feature selection methods and hyper parameter tuning have been implemented in this work leading to a further increase in accuracy. The authors have created an ensemble of various individual classifiers (Support Vector Machine, Decision Trees, K Nearest Neighbors, Random Forest, and Gradient Boosting) Various classification algorithms were analyzed and compared for their performance on the Framingham dataset. Some classifiers have shown good performance while other classifiers have shown poor performance.

The authors in **[10]** introduced a sizable patient database for the study of diabetes. In order to diagnose Type 2 diabetes, the dataset is transformed into a time series and examined using temporal predictive Deep Learning models. The KAIMRCD dataset was used in the study that was just given, and 2 forms of training were used with an accuracy of above 97%. However, due to hardware constraints, the development is sluggish. The surgical workflow recognition problem was significantly improved by the authors in **[11]** by applying a CNN with an LSTM and capitalizing on some information and temporal that saved form system. This study demonstrated that the action still has a lot of restrictions, which results in insufficient datasets.

The global epidemiology of type 2 diabetes was examined by the authors of **[12]** Based on epidemiological data from the Global Burden of Disease (GBD) current dataset from the Institute of Health Metrics, Seattle, they examined the incidence, prevalence, and burden of suffering of diabetes mellitus. To obtain estimates about this condition, they gathered all type 2 diabetes data from 1990 to 2017 and used this data in a Time Series Modeler. Without using AI, this study arrived at its conclusion, making it possible to identify human error.

In **[13]** the author used a support vector machine to categorize patient data before passing the results to a time series analysis to control the patient's medications and find the best medication for them to take. However, the authors' novel technique in their study performed worse than reference techniques because of the datasets they used.

The **[14]** suggests two deep learning models (Attention LSTM-FCN and Long Short Term Memory Fully Convolutional Network, or LSTM-FCN) for classifying multivariate time series. These suggested models are tested on 35 datasets with little to no preprocessing and feature extraction, and they perform well in the majority of them. This study tests the datasets on DTW, Random Forest, SVM with a linear kernel, and SVM with a third degree polynomial kernel to determine baselines, then uses the dataset with the highest score as the baseline.

In order to address the issue of data scarcity, the **[15]** propose a prototype embedding framework called Deep Prototypical Networks (DPN), which uses a main embedding space to capture differences between various time series classes. The paper also adds a relationship-dependent masking module to the DPN framework to automatically combine pertinent data with a distance metric learning process, which solves the problem of data imbalance and provides reliable time series classification.

LSTM-FCN **[16]** is a deep learning framework that converts the LSTMFCN models of UTS into MTS by enhancing them with squeeze-and-excitation blocks. It uses an LSTM layer and stacked CNN layer to extract features for a softmax layer to predict the label for classification.

The **[17]** applied the following ML algorithm techniques: random forest, decision tree, and XGBoost to a bank card fraud data set. The algorithms provided similar results to each other, and they were all highly accurate. The best accuracy went to the XGBoost algorithm.

3 Proposed Methodology

We propose a novel technique to increase the accuracy of detecting diabetes. Our proposed technique focuses on applying some artificial intelligence algorithms such as neural networks and other algorithms in the training and classification stage and comparing these algorithms to find the best and fastest results.

We will use a linear SVM because the features extracted from patients are not large in size and do not need a lot of processing. Furthermore, the linear SVM is fast, and will not consume a lot of time.

Also, because artificial neural networks (ANNs) are reliable statistical techniques that attempt to mimic the functioning of the human nervous system by establishing a logical model made up of interconnected neurons in a computing network, we will use an ANN classifier. Complex modeling problems like estimation, classification, and pattern recognition are solved using neural networks. The network is trained by adjusting the weights between neurons, allowing it to predict output value(s) after receiving a variety of instructive data from prior experiments.

The results of using the XGBoost algorithm have demonstrated that the designed system is highly selective for acetone, even at low concentrations. Moreover, it was demonstrated that XGBoost performs best when compared to other widely used algorithms.

In our research, we focus on following up on giving treatment to the patient and following up on the patient's susceptibility to treatment through a graph showing the patient's susceptibility to treatment.

Many tests are taken for the patient (features) periodically, then these features are entered into the Machine training system and work to extract the results and compare these results with the previous features.

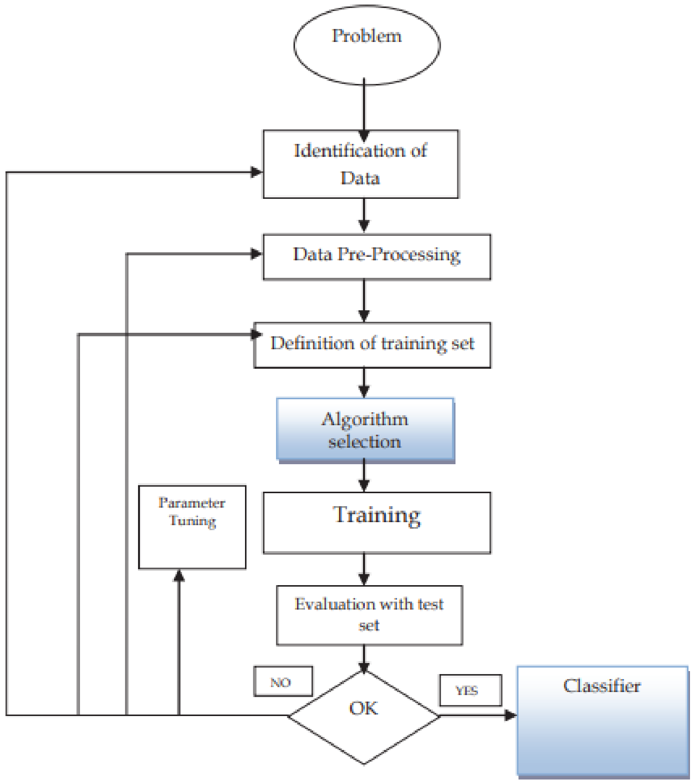

Figure 1: Proposed work

4 Wireless Sensor Network

The WSN contains a huge number of nodes randomly deployed in a hostile environment, connected to one base station. The network is divided into zones, with the nodes within each zone acting as monitors for events indicating the occurrence of a fire, and thus triggering an alarm that is sent to the base station (BS) **[18]**.

A large number of these disposable sensors can be networked in many applications that require unattended operations. A Wireless Sensor Network (WSN) contains hundreds or thousands of these sensor nodes. These sensors have the ability to communicate either with each other or directly with an external base station (BS). A greater number of sensors allows for sensing over larger geographical regions with greater accuracy. Basically, each sensor node comprises a sensing, processing, transmission, mobilizer, position-finding system, and power units (some of these components are optional like the mobilizer). Sensor nodes are usually scattered in a sensor field, which is an area where the sensor nodes are deployed. Sensor nodes coordinate among themselves to produce high-quality information about the physical environment **[19]**.

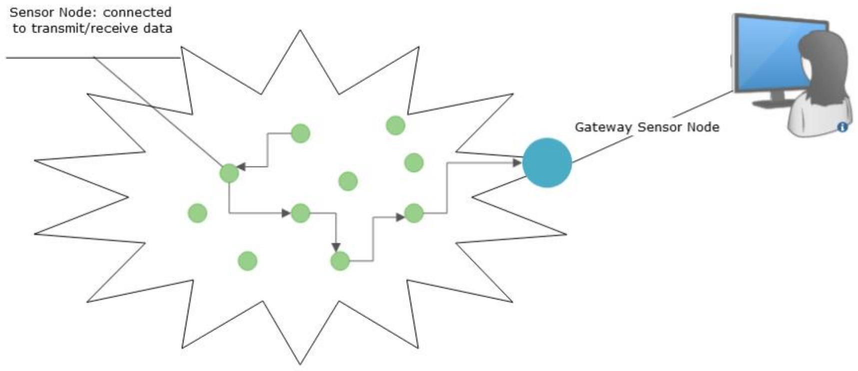

Figure 2:WSN

4.1. The routing protocol:

The routing protocol is a process to select a suitable path for the data to travel from source to destination. The process encounters several difficulties while selecting the route, which depends upon, the type of network, channel characteristics, and performance metrics **[20]**.

The data sensed by the sensor nodes in a wireless sensor network (WSN) is typically forwarded to the base station that connects the sensor network with the other networks (maybe the internet) where the data is collected and analyzed and some action is taken accordingly.

In very small sensor, networks where the base station and motes (sensor nodes) are so close that they can communicate directly with each other this is single-hop communication but, in most WSN applications the coverage area is so large that requires thousands of nodes to be placed and this scenario requires multi-hop communication because most of the sensor nodes are so far from the sink node (gateway) so that they cannot communicate directly with the base station. The single-hop communication is also called direct communication and multi-hop communication is called indirect communication.

In multi-hop communication, the sensor nodes not only produce and deliver their material but also serve as a path for other sensor nodes toward the base station. The process of finding a suitable path from the source node to the destination node is called routing and this is the primary responsibility of the network layer.

The routing protocols define how nodes will communicate with each other and how the information will be disseminated through the network.

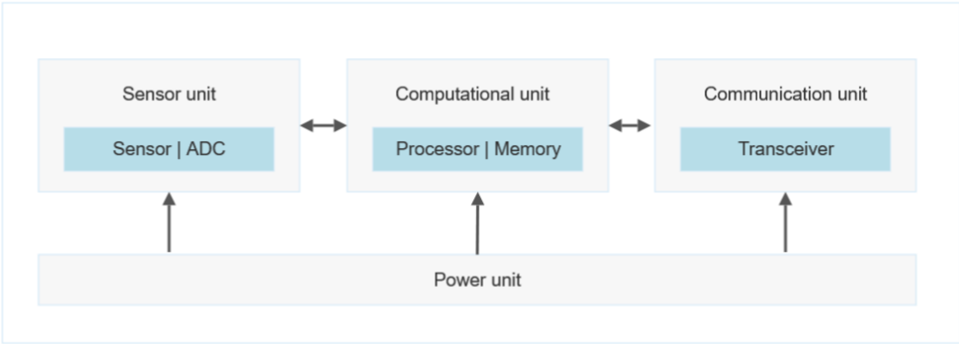

Figure 3:Sensor Architecture

4.2. Transmission Data:

Data transmission between nodes occurs randomly in all areas of the network by using the shortest path protocol and multi-hop protocol **[21]**. Then the last station of the data is a base station, it collects all transmitted information, then analysis this information to make decisions.

The shortest path can be found to connect all the nodes. A sensor network can be regarded as a weighted undirected graph, a node can be regarded as a vertex in the weighted undirected graph, and a connection between any two nodes can be regarded as an edge connecting two vertices, and its weight value is the length of the power line. **[22]**

4.3. Used Techniques for Transmission

a) Transmission time: is the amount of time from the beginning until the end of a message transmission. In the case of a digital message, it is the time from the first bit until the last bit of a message has left the transmitting node.

$$\boldsymbol{TransmissionTime = Packetsize/BitRate}$$

b) TimeTock: is the time of transmission.

c) Power: is the current energy of the node after transmission process, this energy calculated depends on the status of the node, as for example, the transmission mode consume different power than sleep.

4.3.1 Nodes State:

All nodes in the network have many states. In first run of the network the node status is idle, then when a node sends a broadcast message it converted to a send state after sending the message the node converted to receive state to receive an acknowledgment, if the receiver node was busy it converted to a wait state, finally after all steps completed, the node goes to sleep or idle state.

Any transmission process between nodes consumes energy, in WSN the power has limited, this causes we should save the lifetime of the network and transmission data with the lowest energy-consuming.

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L}$$

Where:

$\boldsymbol{P_t}$: Transmission power

$\boldsymbol{d}$:T-R separation distance in meters

$\boldsymbol{P_r(d)}$: Received power

$\boldsymbol{L}$: System loss factor ($\boldsymbol{L \geq 1}$)

$\boldsymbol{G_t}$: Transmitter gain

$\boldsymbol{\lambda}$: Wavelength in meters

$\boldsymbol{G_r}$: Receiver gain

4.3.2 Received Signal Strength Indication (RSSI):

RSSI: the relationship between RSSI values and distance is the foundation and the key of ranging and positioning technologies in wireless sensor networks.

RSSI, TOA **[23]** , TDOA **[24]**,and the AOA **[25]** are ranging technologies commonly used now. Just as the use of RSSI ranging need less communication overhead, lower implementation complexity, and lower cost, so it is very suitable for the nodes in wireless sensor network which have limited power.

The principle of RSSI ranging describes the relationship between transmitted power and received power of wireless signals and the distance among nodes. We can show more details in the following equations:

$$\boldsymbol{P_r = P_t \times \left(\frac{1}{d}\right) n}$$

$$\boldsymbol{10\, log\, P_r = 10\, log\, P_t - 10n\, log\, d}$$

$$\boldsymbol{P_r(dbm) = A - 10n\, log\, d}$$

Where:

$\boldsymbol{P_r}$: received power,

P_t: transmitted power,

d: distance between the sending nodes and receiving nodes,

n: transmission factor

RSSI **[26]** is the technique that uses the spatial correlation between the signal strength and physical location to detect the presence of Sybil attacks. This method measures the received power associated with an incoming message(s) from a node and relates this to a unique location and subsequently a unique ID. Usually, there are several nodes measuring this value in order to triangulate this position. In the event that another message is received having the same location but a claim to a different ID, the system will assume that this entity is a Sybil node. Usually, a minimum of three nodes are required but there are claims that four nodes are required to locate a node using RSSI effectively. The choice of the frequency of RSSI calculations varies for different solutions based on the researcher's goal. Sometimes, this value is calculated only if a new node sends a message or if the system is in running, while other solutions run this calculation on every message.

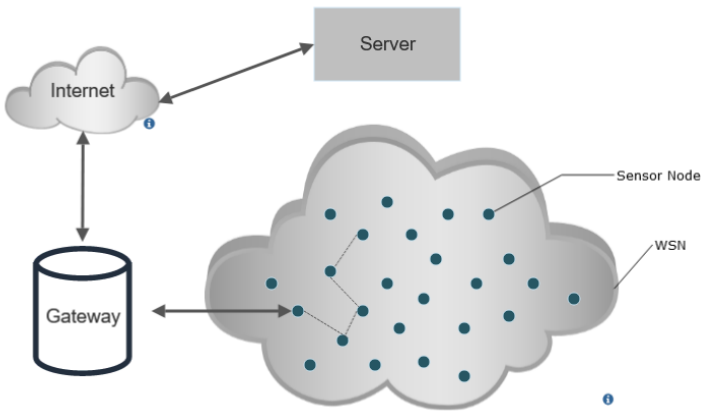

Figure 4:Network Architecture

As shown in the figure above the network architecture contains sensor nodes, WSN, gateway and server

- Sensor nodes: send and receive data in the network.
- WSN: build a flexible area for transmission data.
- Gateway: connect sensor to server through internet.
- Server: all detection processes, and any configuration is done in the server.

5 Results, Analysis and Discussions

In our simulation scenario, we used separate data for each reading feature. We divided it into 70% Training and 30% Test. The features are placed randomly in the simulation using MATLAB simulation. We run the simulation with the initial parameters shown in the following table.

Parameter	Value
Number of Data	769
Number of Training Method	3
Test Data	30%
Training Data	70%
Pregnancies (AVG)	3.845
Glucose (AVG)	120.894
Blood Pressure (AVG)	69.105
Skin Thickness (AVG)	20.536
Insulin (AVG)	79.799
BMI (AVG)	31.992
Diabetes Pedigree Function (AVG)	0.471
Age (AVG)	33

Table 1 Parameters Table

In this paper, we run our simulation using MATLAB on a PC computer core I7 with 16G ram to get the patients result through using SVM, ANN and XGBoost algorithms for classifications, and the final result displayed in the following table:

Table 2 algorithm comparison

Test NO	SVM	ANN	XGBoost
1	91.7%	88.3%	94.4%
2	89.6%	88.3%	94.0%
3	92.6%	88.3%	94.2%
Average	91.3%	88.3%	94.2%

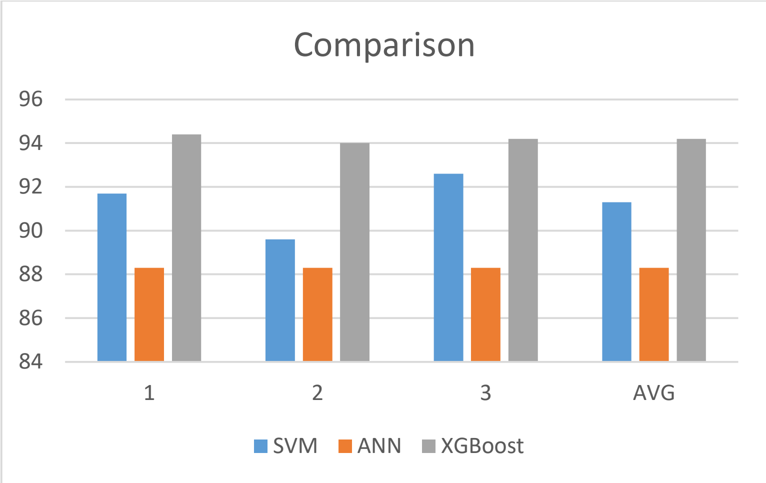

Figure 5 Simulation Result

In Table (2) the accuracy of SVM reached 91.3% because the data can be divided linearly and at the same time it is not considered perfect due to overlapping data. On the other hand, the accuracy of ANN reached 88.3%, which is the lowest accuracy due to the difficult generalization resulting from the small training data set. The best performance was achieved by XGBoost with an accuracy of 94.2% due to its nature of using multiple trees for decision making and its ability to perform parallel computation.

6 Conclusions

Technology never stops evolving and advancing, and this applies to our thesis in which we have implemented artificial intelligence algorithms on the time series of diabetes, which detect whether a patient have diabetes or not by analyzing the time series data.

In this thesis, we studied the best techniques and methods used in machine learning by comparing algorithms (ANN, SVM, XGboost), and XGboost algorithm was the best, and to follow up on the patient's condition after discovering diabetes, we used WSN to take periodic readings to find out the extent of the patient condition improvement after the proposed treatment of The specialist doctor, which enhances the feasibility of developing a treatment for diabetes and its potential future uses.

Among the future work is to add more features and compare them with the current features, and to add another classifier such as K Nearest Neighbor (KNN).

References

[1] H. A. S. و. I. H. J. a. others, "Intelligent Diagnostic Prediction and Classification System for Parkinson's Disease by Incorporating Sperm Swarm Optimization (SSO) and Density- Based Feature Selection Methods," *Int. J. Advance Soft Compu. Appl IJASCA,* vol. 15, no. 1, March 2023.

[2] T. S. a. S. Unhapipat, "Time series analysis of diabetes patients," *A case study of Jigme Dorji Wangchuk National Referral Hospital in Bhutan,* 2018.

[3] T. HamadAlanazia, "Health technology diffusion: Case of remote patient monitoring (RPM) for the care of senior population," *Technology in Society*, vol. 66, 2021.

[4] B. G. G. T. Soltani, "Covert single-hop communication in a wireless network with distributed artificial noise generation". 2014 52nd Annual Allerton Conference on Communication, Control, and Computing," *Annual Allerton Conference on Communication, Control, and Computing (Allerton),* vol. 2, no. 1, p. 1078–1085, 2014.

[5] M. P. Ruth S Weinstock, "Patient education: Type 1 diabetes: Overview (Beyond the Basics)," *uptodate*, Jul 2022.

[6] N.-E. R. Aljawharah Alnasser, "New trust-based detection model for concealed malicious nodes in a wireless sensor network," *ResearchGate,* 2015.

[7] A. V. a. others, "Diabetes Diagnostic Prediction Using Vector Support Machines," *ScienceDirect*, vol. 170, p. 376–381, 2020.

[8] T. S. a. S. Unhapipat, "Time series analysis of diabetes patients," *A case study of Jigme Dorji Wangchuk National Referral Hospital in Bhutan,,* 2018.

[9] D. S. 2. Priyanka Gupta1, "Improving the Prediction of Heart Disease Using Ensemble Learning and Feature Selection," *Int. J. Advance Soft Compu IJISCA,* vol. 14, no. 2, pp. 1-13, July 2022.

[10] Z. A. a. others, "Type-2 Diabetes Mellitus Diagnosis from Time Series Clinical Data using Deep Learning Models," *Research gate,* 2018.

[11] "Workflow Recognition from Surgical Videos using Recurrent Convolutional Network," *IEEE Transactions on Medical Imaging,* p. 1114–1126, 2017.

[12] M. A. B. K. a. others, "Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted Trends," *J Epidemiol Glob Health,* pp. 101-111, 2020.

[13] C. P. N. V. A. T. A. N. Bui, "Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases," *International Conference on the Development of Biomedical Engineering in Vietnam, Springer,* pp. 809-818, 2017.

[14] S. M. a. o. Fazle Karim, "Multivariate LSTM-FCNs for Time Series Classification," *arXiv:1801.04503v2 [cs.LG],* 2019.

[15] C. H. a. others, "Deep Prototypical Networks for Imbalanced Time Series Classification under Data Scarcity," *CIKM*, 2019.

[16] H. I. Fawaz, G. Forestier, J. Weber and L. Idoumghar, "Deep learning for time series classifcation," *Data Mining and Knowledge Discovery,* pp. 1-47, 2019.

[17] M. A. a. A. K. V. Jain, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," *ICRITO IEEE 8th Int. Conf. Reliab,* p. 86–88, 2020.

[18] N.-E. R. a. A. Alnasser, "Lightweight trust model for the detection of concealed malicious nodes in sparse wireless ad hoc networks," *International Journal of Distributed Sensor Networks,* vol. 12, pp. 1-16, 2016.

[19] G. C. a. F. G. L. Y. Fengyun Li, "A Novel Cooperation Mechanism to Enforce Security in Wireless Sensor Networks," *IEEE Fifth Interantion Conference on Genetic and Evolutionary Computing,* Aug. 29 – Sep. 1, 2011.

[20] N. S. a. S. R. Hassan, "Routing Protocols for Wireless Sensor Networks (WSNs)," *intechopen,* no. 10.5772, October 4th 2017.

[21] M. F. Y. ,. K. A. A. Moustafa A. Youssef, "A Constrained Shortest-Path Energy-Aware Routing Algorithm for Wireless Sensor Networks," *IEEE,* 1997.

[22] P. R.-P. Juan Cota-Ruiz, "A Recursive Shortest Path Routing Algorithm With Application for Wireless Sensor Network Localization," *IEEE SENSORS JOURNAL,* vol. 16, no. 11, p. 4631, 2016.

[23] K. Yu and Y. J. Guo, "Non-Line-of-Sight Detection Based on TOA and Signal Strength," *Personal, Indoor and Mobile Radio Communications,* pp. 1-5, 2008.

[24] J. G. L. a. C. G. P. J. I. Kim, "A Mitigation of Line-of-Sight by TDOA Error Modeling in Wireless Communication System," *International Conference on Control, Automation and Systems, Seoul,* 2008.

[25] D. N. a. B. Nath, "“Ad Hoc Positioning System (APS) Using AoA [A],," *Proceedings of the IEEE INFOCOM [C], New York,* pp. 1734-1743, 2013.

[26] ∗. G. M. B. 2. a. R. A. 2. Omar Cheikhrouhou 1, "A Hybrid DV-Hop Algorithm Using RSSI for Localization in Large-Scale Wireless Sensor Networks," *Sensors,* 2018.

[27] C. D. M. T. A. M. O. a. K. S. D. O. Abegunde, "The burden and costs of chronic diseases in low-income and middle-income countries," *Lancet,* vol. 370, no. 9603, p. 1929–1938, 2007.

[28] X. F. a. o. Mingqi Li, "Diabetes Prediction Based on XGBoost Algorithm," *IOP Conference Series: Materials Science and Engineering,* pp. 1-7, 2020.