DiabSense: early diagnosis of non-insulin-dependent diabetes mellitus using smartphone-based human activity recognition and diabetic retinopathy analysis with Graph Neural Network

Md Nuho Ul Alam[1], Ibrahim Hasnine[1], Erfanul Hoque Bahadur[1], Abdul Kadar Muhammad Masum[2*], Mercedes Briones Urbano[3,4,5], Manuel Masias Vergara[3,6,7], Jia Uddin[8], Imran Ashraf[9*] and Md. Abdus Samad[9*]

*Correspondence:
masum.swe@diu.edu.
bd; imranashraf@ynu.ac.kr;
masamad@yu.ac.kr

[1] Department of Computer
Science and Engineering,
International Islamic
University Chittagong, Kumira,
Chittagong 4318, Bangladesh
[2] Department of Software
Engineering, Daffodil
International University,
Dhaka 1216, Bangladesh
[3] Universidad Europea del
Atlantico, Isabel Torres 21,
Santander 39011, Spain
[4] Universidad Internacional
Iberoamericana,
Campeche 24560, Mexico
[5] Universidad de La Romana, La
Romana, República Dominicana
[6] Universidad Internacional
Iberoamericana Arecibo, Puerto
Rico 00613, USA
[7] Fundacion Universitaria
Internacional de Columbia,
Bogota, Colombia
[8] AI and Big Data Department,
Endicott College, Woosong
University, Daejeon 34606, South
Korea
[9] Department of Information
and Communication
Engineering, Yeungnam
University, Gyeongsan-si, South
Korea

Abstract

Non-Insulin-Dependent Diabetes Mellitus (NIDDM) is a chronic health condition caused by high blood sugar levels, and if not treated early, it can lead to serious complications i.e. blindness. Human Activity Recognition (HAR) offers potential for early NIDDM diagnosis, emerging as a key application for HAR technology. This research introduces DiabSense, a state-of-the-art smartphone-dependent system for early staging of NIDDM. DiabSense incorporates HAR and Diabetic Retinopathy (DR) upon leveraging the power of two different Graph Neural Networks (GNN). HAR uses a comprehensive array of 23 human activities resembling Diabetes symptoms, and DR is a prevalent complication of NIDDM. Graph Attention Network (GAT) in HAR achieved 98.32% accuracy on sensor data, while Graph Convolutional Network (GCN) in the Aptos 2019 dataset scored 84.48%, surpassing other state-of-the-art models. The trained GCN analyzed retinal images of four experimental human subjects for DR report generation, and GAT generated their average duration of daily activities over 30 days. The daily activities in non-diabetic periods of diabetic patients were measured and compared with the daily activities of the experimental subjects, which helped generate risk factors. Fusing risk factors with DR conditions enabled early diagnosis recommendations for the experimental subjects despite the absence of any apparent symptoms. The comparison of DiabSense system outcome with clinical diagnosis reports in the experimental subjects was conducted using the A1C test. The test results confirmed the accurate assessment of early diagnosis requirements for experimental subjects by the system. Overall, DiabSense exhibits significant potential for ensuring early NIDDM treatment, improving millions of lives worldwide.

Keywords: Graph Neural Network, Diabetic retinopathy, Human activity recognition, Diabetes, NIDDM

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 2 of 37

Introduction

Diabetes is recognized as a rapidly expanding worldwide health crisis in the twentyfirst century [1]. The number of people affected by this condition in 2021 had surpassed half a billion worldwide (10.5% of the global population). According to projections, this figure will increase to 12. 2% by 2045. Type 2 Diabetes/Non-Insulin-Dependent Diabetes Mellitus (NIDDM) stands as the predominant form of Diabetes, constituting more than 90% of Diabetes cases globally. Another cause for concern is that around 45% of diabetic patients worldwide are currently living with undiagnosed Diabetes, predominantly NIDDM. Early diagnosis and treatment can significantly reduce or prevent the severity of the disease and its associated complications. However, one-third to one-half of individuals with type 2 Diabetes remain undiagnosed due to the absence of apparent symptoms in the early stages. The precise timing of type 2 Diabetes onset is often difficult to ascertain, resulting in a prolonged pre-diagnostic phase. If the diagnosis is significantly delayed, complications can arise, eventually leading to the detection of the condition [2, 3]. These statistics and medical concerns underscore the urgent need to recognize the early signs as well as improve the ability to diagnose people with Diabetes, many of whom are unaware they have the disease, and provide appropriate and timely care for all individuals before it gets too severe.

People with NIDDM are at a higher risk of experiencing various medical complications. Among these complications, including retinopathy, cardiovascular disease, neuropathy, and nephropathy, retinopathy is the most common complication of diabetes [4, 5]. Diabetic Retinopathy (DR) is an eye disorder that can lead to vision loss in persons with diabetes. The existence of DR suggests that the NIDDM has already affected the microcirculation, hence it can be regarded as a valid biomarker of the detrimental consequences of NIDDM in a given individual.

Human Activity Recognition (HAR) is a system that utilizes sensor data to identify activities based on body motions. NIDDM has a negative correlation with physical activity. Sedentary behavior (such as lying down, standing up, sitting) and a lack of physical exercise are associated with higher health risks for NIDDM [6, 7]. The HAR system can detect various sedentary behaviors and physical activities of humans. Studies have shown that individuals who are more physically active have a significantly lower risk of NIDDM than those who are less active [6–9]. Therefore, the HAR system can be valuable for actively classifying human activities over a long period to generate risk factors of NIDDM.

This work aims to improve early diagnosis of NIDDM so that there is no longer a need for humans to depend on apparent medical symptoms. The main objective of DiabSense is to provide individuals with an understanding of their likelihood of having diabetes by tracking their daily activity patterns as well as grading retinal fundus images. DiabSense notifies them about the severity of NIDDM using only their smartphone sensors. This way the system will encourage them to undergo clinical testing procedures based on their risk factors, reducing the reliance on costly routine medical checkups until the system signals an early diagnosis. Early diagnosis facilitated by the proposed system can help prevent or delay the onset of complications related to NIDDM. It is in fact, regular screening is key to preventing blindness caused by Diabetic Retinopathy (DR). Being the

sole system reliant on smartphones, it presents a unique and optimistic approach within the field of medical science.

In this work, a novel system "DiabSense" is presented for the early staging of NIDDM after being motivated by the correlation between daily activities and NIDDM, coupled with Diabetic Retinopathy (DR), the most prevalent complication of NIDDM. Activities that help identify symptoms related to NIDDM were recognized as useful data from smartphone sensor. The smartphone triaxial accelerometer sensor data were utilized to train a Graph Attention Network (GAT) for the Human Activity Recognition (HAR) task. Besides, Vision GNN (ViG), a Graph Convolutional Network (GCN) based model was trained on a benchmark retinal fundus dataset Aptos 2019 to grade DR [10]. Since the DR lesions are usually not quadrate but irregular-shaped, the conventional approach of using CNN-based architecture would not be ideal in this case. CNN treats images as grid structures, which limits their ability to handle complex and irregular shape lesions in fundus images alone. To overcome this limitation and explore a novel vision algorithm, this work delved into the usage of GCN with the ViG algorithm for DR grading. The ensemble of GCN-based ViG and CNN-based EffNet takes into account both irregular and regular shapes of the objects from images which enhances the performance. For experimental purposes and initial validation, human experimental subjects (ES) were chosen to volunteer. Android smartphone device was used to gather sensor and retinal image data from the experimental subjects. Over thirty consecutive days, labelless sensor data on the experimental subjects day-to-day activities were collected. The gathered sensor data and retinal images from the experimental subjects were then fed into the pre-trained GAT and ViG, respectively. Through predictions, the models were able to recognize daily activity patterns and grade their DR condition. By analyzing the predicted activity log, the average duration spent on each activity daily by the experimental subjects was calculated. Additionally, a comprehensive survey of activity patterns exhibited by 97 diabetic patients before their diabetic period lifestyle was conducted. A comparison of the activity patterns of the experimental individuals and diabetes patients was used to determine the risk factor. The six basic physical characteristics-height, weight, blood pressure, age, gender, and evidence of diabetes in first-degree relatives as well as the exercise patterns are taken into consideration in order to assess this resemblance. The calculated Cosine similarity measures for four experimental subjects categorized them according to their risk factors of NIDDM. Furthermore, to validate the risk factor diagnosis, the experimental subjects underwent clinical testing using the A1C diagnostic process, which measures the level of A1C in the blood. The A1C levels of experimental subjects confirmed the generated risk factors by the system were valid. This clinical confirmation aligns with the conclusion reached through the initial assessment of this work, reinforcing the validity of the system.

There are several contributions that this study makes. First and foremost, the main contribution is presenting the first-ever smartphone-based system for the early staging of NIDDM by incorporating DR and HAR with Graph Neural Networks (GNN). Second, time-series data for 23 activities that help identify symptoms of NIDDM were collected using tri-axial smartphone sensors in four possible pocket orientations from 12 volunteers for the HAR phase of this system. To obtain more detailed data, different activities were classified into three levels: normal, moderate, and vigorous, taking

into mind that energy loss correlates to activity intensity. To improve system reliability, data collection was extended to include both indoor and outdoor settings. Third, using GAT to capture internal connections in time-series tri-axial smartphone sensor data is novel as opposed to RNN/transformer for modeling temporal dependence. GAT is computationally efficient, eliminates costly matrix operations, may act on all nodes at the same time, and easily manages varying node significance while operating on varied neighborhood sizes without prior knowledge of the entire graph structure. Fourth, activities irrelevant to NIDDM were also identified to prevent false recognition and increase system robustness. Fifth, as DR lesions are hardly square but rather irregularly shaped, the traditional technique of applying CNN-based architecture would be ineffective in this scenario. CNN analyzes images as grid structures, limiting its capacity to deal with complex and irregularly shaped lesions in fundus images alone. To overcome the limitations of employing typical CNN-based models for DR grading, GCN was used in this study with tunning hyperparameters. The novel ensemble of GCN-based ViG and CNN-based EfficientNet-B5 models considered both irregular and regularly shaped lesions in retinal fundus images, outperforming earlier state-of-the-art works. Sixth, the results of the whole system from four experimental subjects were verified and validated by the A1C medical testing reports of the experimental subjects. Seventh, to strike a balance between accuracy and energy efficiency, this work adopted a low frequency of 1 Hz for the HAR process, ensuring steady accuracy while minimizing energy consumption. Notably, prior researchers frequently collected sensor data at a frequency of 50 Hz, which used plenty of energy.

The main contributions of this work are summarized as follows:

- Introduced the first-ever smartphone-based system for early staging of NIDDM by integrating DR and HAR with Graph Neural Networks (GNN).
- Collected time-series data for 23 activities using tri-axial smartphone sensors in four pocket orientations from 12 volunteers.
- Activities are classified into three levels: normal, moderate, and vigorous, considering the correlation between energy loss and activity intensity.
- Used GAT to capture internal connections in time-series data, offering computational efficiency and managing varied node significance without prior knowledge of the graph structure.
- Identified and excluded activities irrelevant to NIDDM to prevent false recognition and improve robustness.
- Addressed the limitations of CNNs in handling irregularly shaped DR lesions by using a novel ensemble of GCN-based ViG and CNN-based EfficientNet-B5 models.
- Validated the system's results with A1C medical testing reports from four experimental subjects.
- Adopted a low frequency of 1 Hz for the HAR process to balance accuracy and energy consumption.

The rest of this article is organized as follows. "Introduction" section of this article introduces the motivation and necessity of this study. "Literature review" section cover in-depth reviews of the literature that will guide the readers further into the areas of

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 5 of 37

research. "Datasets description and Proposed system" sections include a description of the dataset used and the proposed methodologies. "Results and discussion" section contains the results and discussion and "Conclusion" section presents the conclusions of this study.

Literature review

It is widely known that sedentary behavior (e.g. such as lying down, standing up, sitting) and a lack of physical exercise are connected with higher health risks for NIDDM [6, 7]. General exercise can serve as a preventive measure against NIDDM, and more intensive physical activities are probably even more beneficial than less intensive activities. Colberg et al. [8] emphasized the benefits of long-term periodic activities, provided that these maintain the necessary compliance levels or prevent hyperglycemia lowering triglyceride-rich very low-density lipoproteins (VLDL) readings. An eight-year follow-up on a cross-sectional study of 87,253 American women aged 34 to 59 found that the risk factor for NIDDM was two-thirds lower in the most active women than in the least active women [9]. These studies suggest that the most involved persons with sedentary behavior and physical activities had a two-thirds reduced risk of NIDDM than those who were least actively engaged. Hence, DiabSense took into account various sedentary behaviors and physical activities for its Human Activity Recognition (HAR) module to generate NIDDM risk factors.

Sedentary behaviors and physical activities can be detected with HAR systems. HAR has been a subject of interest since the 1980 s, owing to its applicability in diverse fields like safety, medicine, sports, robotics, HCI, elderly care, behavioral biometrics, and more [11]. Retrospectively, various sensors data (e.g. accelerometer, gyroscope, temperature, relative humidity, camera) were used by scholars to recognize physical activities in the HAR framework [12–16]. Some studies employed camera and ambient sensors to capture photos or video to execute HAR [17–20]. Few of the studies used wearable sensors to convert human motion into signal patterns for HAR [21–25]. Smartphone sensors were also widely utilized in studies for the incorporation of HAR [26–28]. Guidoux et al. [29] proposed a technique for activity prediction based on embedded sensors in smartphones to quantify energy expenditure while identifying spontaneous physical activities. In some instances, researchers have focused on individual sensors for activity detection (e.g. using only an accelerometer sensor) [30–33]. Sensor-driven systems rely on user-technology collaboration, requiring a balance between user rights and system efficiency. Cameras are the most popular for HAR among all other sensors since the users are visible in the display. However, such systems can raise privacy concerns for users. DiabSense tackles this issue by using only a smartphone accelerometer sensor. Wearable and embedded smartphone sensor-based approaches for activity recognition systems typically avoid user privacy issues. Furthermore, unlike wireless signals-based methods, smartphones are not location-dependent and do not pose any health risks due to radiation.

Researchers have adopted various deep learning algorithms with mobile sensor data to implement HAR due to its suitability for field research [22]. Ignatov [31] identified six activities using tri-axial accelerometer sensor data from the WISDM and UCI datasets with their proposed CNN architecture. The author got 93.32% overall accuracy

on WISDM and 97.62% on the UCI dataset. The researchers Alsheikh et al. [34] constructed a DBN along with various conventional and widely used classifiers with accelerometer sensor data from benchmark Skoda, Daphnet, and WISDM datasets, and got an accuracy of 89.38%, 91.5%, and 98.23% respectively. In their study, Ha and Choi [30] employed a multimodal CNN that utilized two-dimensional kernels. They applied this approach to both the M-health and Skoda datasets, achieving impressive accuracy rates of 98.26% and 97.92% respectively. Xia et al. [35] achieved 95.78% in UCI-HAR, 95.85% in WISDM and 92.63% in OPPORTUNITY dataset by applying LSTM-CNN. LSTM-CNN was first trained on their dataset collected from different mobile sensors. Another study determined that kNN was the best classifier [36]. They applied kNN on the accelerometer and gyroscope sensors collected from the iPod. The researchers demonstrated accuracy in several individual activities but failed to successfully distinguish very similar activities. Uddin and Soylu [25] introduced a HAR system based on LSTM-based Neural Structured Learning (NSL). They trained NSL on data from wearable body sensors and achieved 99% accuracy on the MHEALTH dataset. Bao and Intille [37] from MIT are notable for having the most cited work in HAR. They collected data from five biaxial wireless accelerometers positioned across various body regions and achieved over 80% accuracy with a Decision Tree classifier. According to their findings, wireless or mobile embedded accelerometers placed on the thigh were highly successful in activity recognition. It is clear from the reviews that recent advances in HAR systems owe much to efficient deep learning algorithms featuring numerous layers and parameters in the millions. However, these algorithms also exhibit weaknesses. While LSTM is widely used for modeling time-sequential events, it faces challenges like the vanishing and exploding gradient problem hindering long-term information processing. On the other hand, CNNs designed for grid-like data such as images, are less adept at handling sequential structures like time series data. Additionally, the computational complexity of onboard training for DBN in mobile and wearable sensors arises from an extensive parameter initialization process. The graph attentional layer used throughout Graph Attention Network (GAT) comes to the rescue since it is computationally efficient, avoids costly matrix operations, can operate on all nodes simultaneously, and easily handles different node importance while operating on different neighborhood sizes without knowing the whole graph structure beforehand [38]. Therefore, DiabSense incorporated GAT into the Human Activity Recognition (HAR) segment of the framework.

HAR with sensor data-based systems have received more attention in recent years, particularly when it comes to tracking the health or diseases of a person since the conditions are linked to the patterns observed in daily activities [39]. For instance, Umbricht et al. [40] worked on Schizophrenia patients using HAR. A multi-level feature fusion approach for multimodal HAR in smart healthcare applications was presented by Islam et al. [41]. Papadopoulos et al. [42] worked on Parkinson's disease detection with IMU sensor and multi-modal dataset. The researchers Preece et al. [43] proposed HAR systems as a means of connecting prevalent illnesses to people's levels of physical activity. The authors also examined their systems with everyday activity patterns that aided in the treatment and detection of neurological illnesses.

A significant gap in prior research lies in the absence of HAR being utilized for early NIDDM diagnosis, particularly in the case of DR not being integrated with HAR.

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 7 of 37

DiabSense addresses this gap by incorporating HAR into the framework. DR stands out as the most common complication of NIDDM besides cardiovascular diseases, gastritis, nerve and kidney damage, etc [44].

NIDDM patients tend to skip routine eye screening due to time constraints, lack of symptoms, and limited access to specialists [45]. Therefore the risk of developing DR gets high. One effort to overcome this is the use of artificial intelligence (AI) approaches for DR detection and diagnosis. Gulshan et al. [46] proposed an Inception-v3 network trained on 0.13 million images evaluated by 54 U.S. board-certified ophthalmologists. The model, tested on two different datasets classified by 7 U.S. board-certified ophthalmologists and achieved an AUC of 0.97−0.99 for detecting referable DR. Gulshan et al. [47] further validated the performance of the DR grading system across two sites in India compared to manual grading. On the Messidor-1 and Aptos 2019 datasets, Gangwar and Ravi [48] used a pre-trained Inception-ResNet-v2 and added a custom layer on top to obtain accuracies of 72.33% and 82.18%, respectively. Kassani et al. [49] developed Xception architecture with the integration of CNN layers and showed 83.09% accuracy their model on the Aptos 2019 dataset outperformed other CNN-based pre-trained models. Bodapati et al. [50] developed a DNN which they trained with blended multimodal deep features and got around 81% accuracy on the Aptos 2019 dataset. The features were extracted using several pre-trained CNN architectures. Following the same feature extraction and blending procedure, Bodapati et al. [51] later trained a composite gated attention DNN which performed slightly better than before with an accuracy of 82.5%. Adem [52] showed segmenting optic disc region from fundus image before training a CNN gives better results than CNN-only methods alone. Evidently, Convolutional Neural Networks (CNN) are the most used in image processing and computer vision applications. The conventional approaches of using convolutional neural networks and transformers treat images as grid or sequence structures, which may not be ideal for capturing irregular and complex objects. To overcome this limitation, DiabSense delved into Graph Convolutional Network (GCN) for the DR grading module as it takes into account irregularly structured objects in images by extracting graph-level features [53].

Notably, prior researchers often operated with a sensor data collection frequency of 50 Hz, which consumed significant energy [54, 55]. To strike a balance between accuracy and energy efficiency, this work adopted a low frequency of 1 Hz for the HAR process, ensuring steady accuracy while minimizing energy consumption

This paper holds significant importance due to its innovative approach to addressing the early diagnosis of NIDDM, which is a pressing health concern worldwide. The prior studies discussed above have underscored a notable gap by demonstrating the absence of utilization of HAR and DR grading for early NIDDM diagnosis. Moreover, previous researchers have worked on DR and smartphone-based HAR in the healthcare domain separately [48, 51, 56−64], but there was an absence of HAR and DR being jointly utilized for early NIDDM diagnosis. This work addresses this gap by incorporating both into the same framework. Current methods for diagnosing NIDDM involve invasive and expensive tests, such as blood glucose tests, oral glucose tolerance tests, and A1C tests. These tests are often only performed when patients exhibit symptoms of the disease, which may be too late for effective treatment. The DiabSense system introduced in this paper addresses this challenge by

providing a non-invasive, cost-effective method for early NIDDM diagnosis, reducing the reliance on costly routine medical checkups until the system detects early signs of NIDDM. The system encourages people to undergo clinical testing procedures using only their smartphone sensors. This also reduces the need for expensive routine medical examinations. Furthermore, there is no longer a need for people to rely solely on apparent symptoms.

The DiabSense system method was chosen for several reasons based on prior research findings and methodological considerations. Firstly, previous studies have established a correlation between human activities and NIDDM risk factors. To leverage this connection, the DiabSense system integrated the HAR module into the system to generate NIDDM risk factors. Secondly, concerns regarding user privacy and health risks associated with conventional sensor-based systems, such as camera sensor-based systems and wireless signal-based radiations, have been highlighted in earlier research works. To address these concerns, this study opted to utilize smartphone sensors for the data collection process in the HAR phase of the system. By leveraging smartphone sensors, the study aimed to mitigate privacy risks and health concerns while still effectively capturing relevant data for analysis. Moreover, traditional Long Short-Term Memory (LSTM)-based approaches have exhibited computational inefficiencies, particularly concerning the vanishing gradient problem, which hampers long-term information processing. Consequently, the GAT model was selected for the HAR phase of this system due to its computational efficiency, ability to avoid costly matrix operations, simultaneous operation on all nodes, and capacity to handle different node importance levels without necessitating prior knowledge of the entire graph structure. Furthermore, prior research has established DR as the most common complication of NIDDM, serving as a valid biomarker of the detrimental consequences of NIDDM in individuals. To enhance the system's robustness in the final decision-making process, this system classified DR grades and fused the grade report with risk factors generated from HAR. GCN-based ViG model was employed for the DR grading task, addressing the limitations of traditional CNN-based architectures in handling irregularly shaped DR lesions. Overall, by integrating HAR and DR grading into the framework and employing advanced models like GAT and GCN-based ViG, the proposed system aimed to effectively predict early diagnosis while addressing privacy, health, and computational efficiency concerns highlighted in prior research works.

Datasets description

Retinal fundus image

The benchmark Aptos 2019 [10] dataset was used for DR grading. The dataset includes a large set of retinal images captured from rural areas by Aravind Eye Hospital technicians utilizing fundus photography under various imaging conditions. A clinician graded each image on a scale of 0 to 4 for the severity of DR. Since the dataset was gathered via a Kaggle competition, their corresponding test images were kept private. As a result, the 3,662 training images were used. The number of fundus images and representative samples from each class is shown in Table 1 and Fig. 1 respectively.
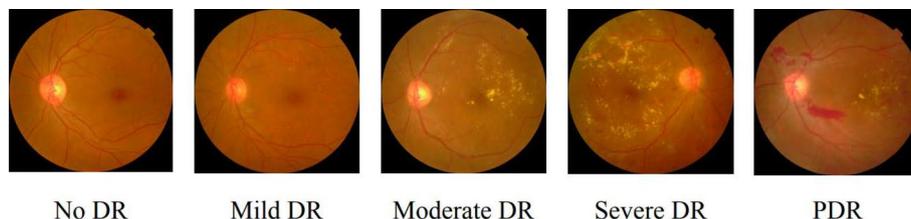
No DR   Mild DR   Moderate DR   Severe DR   PDR

Fig. 1 Samples of each class labels from Aptos 2019 dataset

Table 1 Number of fundus images per class for Aptos 2019 dataset

Severity	Label	No of samples	Ratio in %
0	No DR	1805	49.29
1	Mild DR	370	10.10
2	Moderate DR	999	27.28
3	Severe DR	193	5.27
4	PDR	295	8.06
	Total	3662	

Sensor data and targeted activities

Prominent scholars in the field of HAR have incorporated accelerometer sensors into their research endeavors, yielding commendable outcomes [31, 34, 37]. In alignment with this trend, the research integrates accelerometer sensors, anticipating valuable insights and outcomes. At a 1 Hz frequency, we have collected smartphone-embedded tri-axial accelerometer sensor data of 23 activities through a smartphone app. A tri-axial accelerometer is a sensor that provides estimates of velocity and displacement in addition to acceleration along the x, y, and z axes.

In the context of HAR aimed at assessing Diabetes risk factors, we have selected physical activities closely linked to diabetic symptoms. The selected 23 activities were: Walking-normal (A1), Walking-moderate (A2), Walking-vigorous (A3), Standing (A4), Walking upstairs-normal (A5), Walking upstairs-moderate (A6), Walking upstairs-vigorous (A7), Walking Downstairs-normal (A8), Walking Downstairs-moderate (A9), Walking Downstairs-vigorous (A10), Drinking (A11), Eating (A12), Itching (A13), Lying (A14), Using toilet (A15), Jogging (A16), Cycling (A17), Irrelevant activities (A18), Driving (A19), Sitting (A20), Exercise-dips (A21), Exercise-leg raise (A22), Push up (A23).

The 23 activities are related to the diverse nature of diabetic conditions. Activities associated with cardiovascular movements e.g. walking, jogging, cycling, aerobic exercise e.g. push-ups, and strength training exercises e.g. leg raises were linked to blood glucose control complexities. Recording the urination log involved recognizing activities like drinking and eating, and identifying physical weakness was facilitated by acquisition activities like falling. The ability to identify actions such as reclining, sitting, itching the genitals, and standing improved the capacity to identify symptoms associated with type 2 Diabetes. In pursuit of more specific data, taking into account that energy loss depends on activity intensity, several activities were categorized into three levels: Normal, Moderate, and Vigorous.

   The process of collecting data began with the help of 12 volunteers. Half of them were young, while the other half were aged 50 or above. All participants were in good health without significant physical issues. The dataset used in the trial consisted of around 7,500 instances for each diabetic symptomatic activity, excluding the activities that were deemed irrelevant. Irrelevant activities totaled 36,000 occurrences. Sensor data for irrelevant activities were also collected to avoid false positives (incorrectly accepting that a smartphone is not being used during human activity) or false negatives (incorrectly rejecting the use of a smartphone during any activity). Smartphones may be temporarily placed on surfaces or utilized for various reasons. Users might engage in activities such as browsing, typing, or talking, whether seated, standing, or lying down, all categorized separately as irrelevant activities. A significant concern requiring resolution was the capability of a person to place the smartphone in any pocket position, be it flipped, upside down, or downside up. Addressing this, data was collected with the phone positioned in four possible pocket orientations. To enhance system robustness, data collection extended across both indoor and outdoor scenarios. Subjects engaged in walking activities in diverse locations, encompassing malls, both free and busy roads, in rooms, with sensors capturing walking data. Sitting modes, including squatting and normal sitting for toilet use, were also examined, and sensors recorded data on both lower and higher commodes. In total, the dataset comprises 201,000 occurrences across 23 activities.

Proposed system

This section details the process of building and executing the DiabSense system. Figure 2 provides all the necessary steps followed.
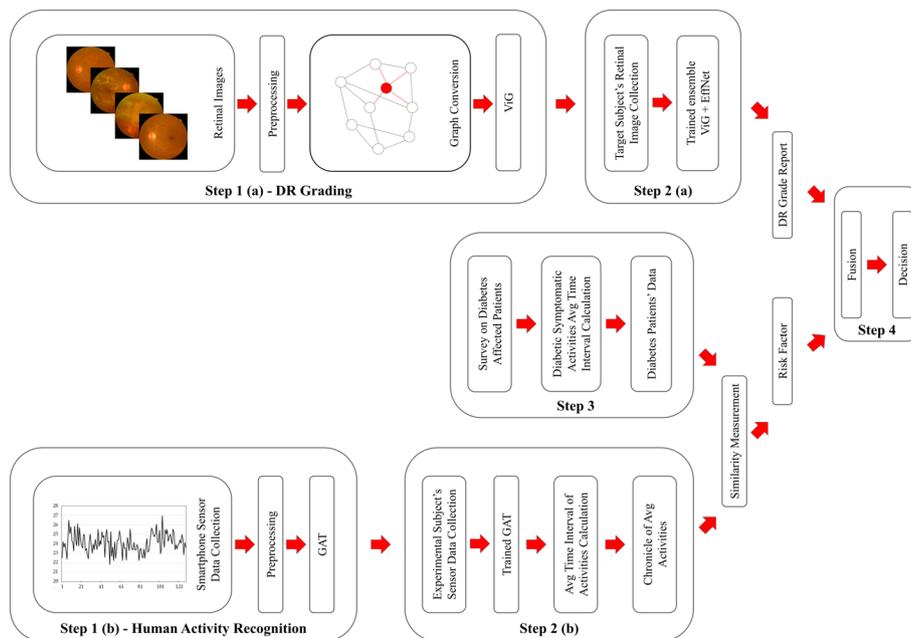

Fig. 2 The DiabSense system architecture

Architecture description

In the initial phase depicted in Fig. 2—Step 1 (a) DR Grading, various preprocessing techniques were applied to the Aptos 2019 retinal fundus dataset [10]. Later, in the graph conversion part, every image got converted into 16x16 size patches. Treating each patch as a node by connecting neighboring patches every image was transformed into graphs which eventually served as input for training ViG.

Moving to Step 1 (b), the Human Activity Recognition phase involved gathering sensor data capturing both symptomatic and asymptomatic activities of diabetic individuals via smartphones. This data was collected from a group of 12 volunteers who willingly participated in the study. Preprocessing techniques were applied before feeding sensor data into the GAT model.

Next, experimental subjects were evaluated for their risk for Diabetes and they provided labelless retinal pictures and 30 days of sensor data, which were then symbolically represented using the results. Four target experimental subjects initially provided labelless retinal images in the first stage of Fig. 2—Step 2 (a). The images were then classified using the trained ViG model. Notably, a soft voting ensemble of ViG and EfficientNet-B5 showed improved performance on the Aptos 2019 dataset, leading to its adoption for classifying the experimental subjects' images. Resulting DR grade reports were generated from these classifications.

In Step 2 (b), 30 days of sensor data from the four experimental subjects were collected and classified using the GAT model pre-trained on volunteers' smartphone sensor data. This facilitated the calculation of average activity time intervals for each subject, contributing to the creation of the chronicle of average activities.

Step 3 involved surveying the Sylhet Diabetic Hospital to gather data from diabetic patients. From a sample of 97 diabetic patients, their primary daily activities average time intervals along with secondary biological data were gathered. Their dataset was finalized as "Diabetes Patients' Data".

In the similarity measurement phase, cosine similarity was employed to measure the similarity between the activity chronicles of experimental subjects and the data from diabetic patients. This similarity measurement aided in assessing the diabetes risk factor for each subject. Later, their risk factor and DR grade report fused to generate a final robust decision regarding their early diagnosis requirements. All architecture components have been thoroughly detailed in the subsequent sections.

Diabetic retinopathy (DR) grading

The first step of the DiabSense system architecture involves grading DR from retinal fundus images. The nature of this problem is challenging and noisy. The assessments provided by doctors, when accessible, exhibit significant variation. According to Fig. 3, each row represents patient images and each column is US-board certified ophthalmologist graders. It is observed that one doctor may classify an eye as severity level 3, while another doctor may classify it as level 1. Interestingly, certain doctors may even assign a severity level of 4! Supporting evidence from the Tensorflow summit was presented by Google [65].
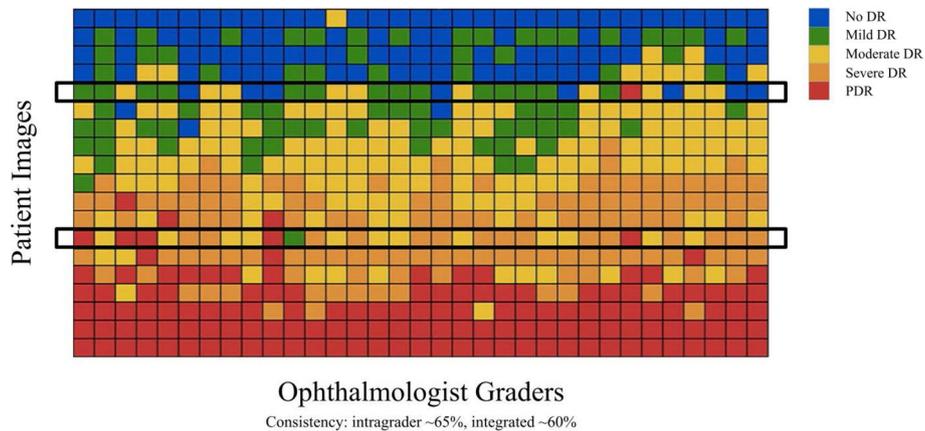
Fig. 3 Inconsistency among ophthalmologists. X-axis represents ophthalmologist graders and Y-axis represents patient images [65]
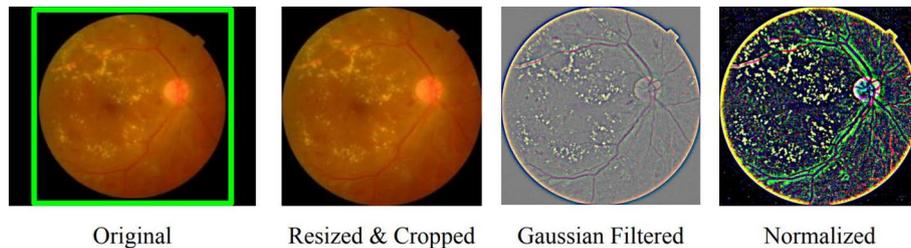

Fig. 4 Preprocessings of Aptos 2019 dataset

Pre-processing fundus Images

80% of the Aptos 2019 data was applied for training, while the rest was used for testing. Resizing and cropping uninformative areas of the fundus images were conducted to prevent model overfitting and erroneous pattern learning. Images larger than 1024px were adjusted to that size while maintaining their aspect ratio. This maintains the original object composition which is crucial for lesion detection. The dataset images had inconsistent black spaces around the retina which may impact the model learning (sample evidence in Fig. 4—Original). This was addressed by cropping the spherical z-space based on the circle radius and center of the fundus images (Fig. 4—Resized and Cropped).

One issue that remained with the images was uneven brightness which was causing some images to look very dark and it was difficult to visualize the lesions as the retinal images were captured under varied imaging conditions. To enhance the performance of ViG and distinguish lesions like hemorrhages, hard exudates, cotton wool patches, aneurysms, and abnormal blood vessel growth, Gaussian filter was applied to equalize brightness. See a sample result in Fig. 4—Gaussian Filtered.
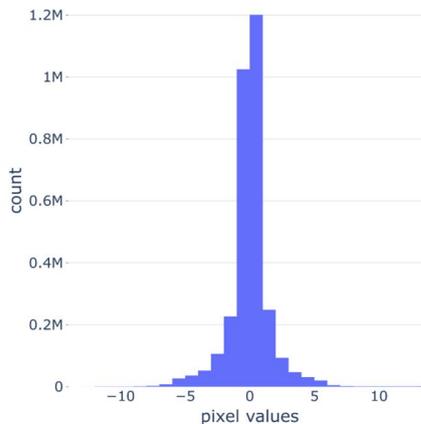
When an RGB image is loaded in memory, the pixel values range from 0 to 255 as 8-bit integers for each of the three channels. However, neural network models prefer floating-point values within a smaller range. Hence, the images were normalized using Z-Score normalization to reduce skewness and enhance the training stability of the training model.

Table 2 Normalized channel mean and standard deviation

Channel	Mean (μ)	Standard deviation (σ)
Red	0.5023	0.0985
Green	0.5019	0.0813
Blue	0.5018	0.0379
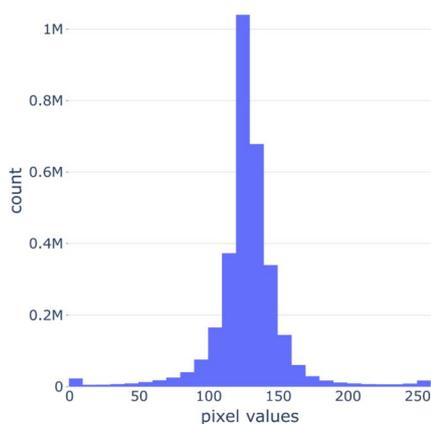
Before z-score normalization   After z-score normalization

Fig. 5 Pixels distribution of the image in Fig. 4—Normalized

This technique was useful as the dataset did not have extreme outliers that needed clipping. Initially, channel-wise mean and standard deviation (see Table 2) were calculated. The calculation began by taking the dataset images in a Python programming language list. Then initialization of accumulators was done to store the sums and sums of squares for each channel across all images in the dataset. The storing process was done by iterating through each image, separating the red, green, and blue pixel values, and updating the sums and sums of squares accordingly for reach channels. After processing all images by iteration, the mean for each channel was calculated by dividing the accumulated sum by the total number of pixels. Then, the variance is found by dividing the accumulated sum by the total number of pixels and then subtracting the square of the mean from this result. The standard deviation is then obtained by taking the square root of the variance. The mean and standard deviation were then utilized in the z-score normalization equation to achieve normalization in the range of -10 to 10, centered around 0, promoting faster convergence (refer to Fig. 5). A sample normalized image is shown in Fig. 4—Normalized.

The Aptos 2019 dataset displayed class imbalance with significant variations in sample numbers across classes. This could potentially result in a biased model. To attain more balanced class distributions within training batches, an oversampling approach was employed. Initially lacking sufficient images from the minority class, training batches in Fig. 6 achieved more balanced class distributions in Fig. 7 after oversampling. Figure 8 shows average representation across all batches was nearly identical
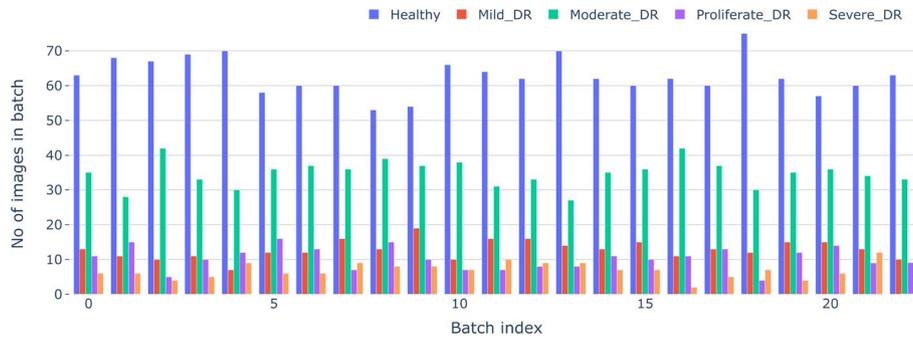
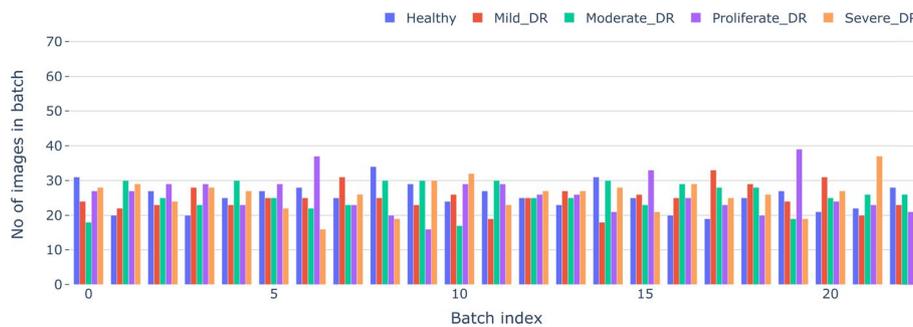**Fig. 6** Training datasets class distribution per batch before sampling

Fig. 7 Training datasets class distribution per batch after sampling
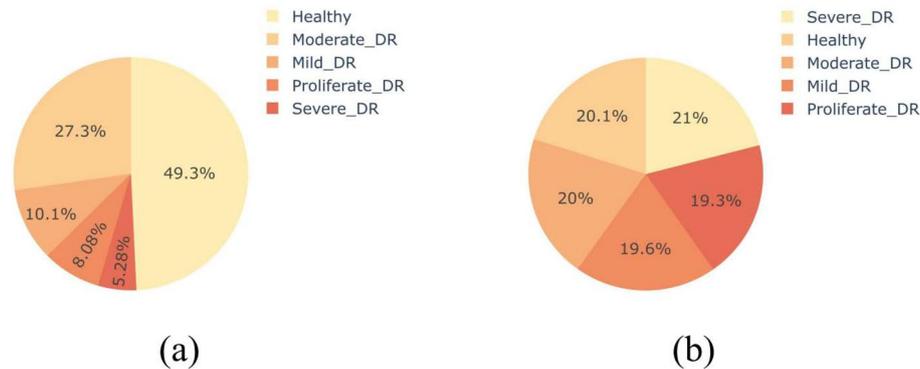
Fig. 8 Average representation of images per class across all batches (**a**) before and (**b**) after sampling

which ensured the model received a consistent proportion of data from each class during training.

Augmentations (follow Fig. 9) were applied to make the model robust from noisy data and promote better generalization by diversifying the training dataset. Overfitting happens when a model becomes highly specialized in collecting patterns from noisy data, resulting in high variance, and poor generalization to unknown samples. Image augmentation tackles this by introducing variations to the training dataset, which prevents the model from memorizing specific details and promotes improved

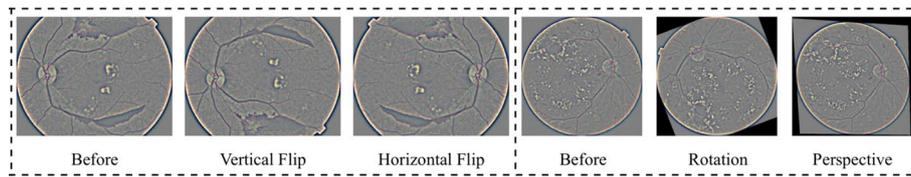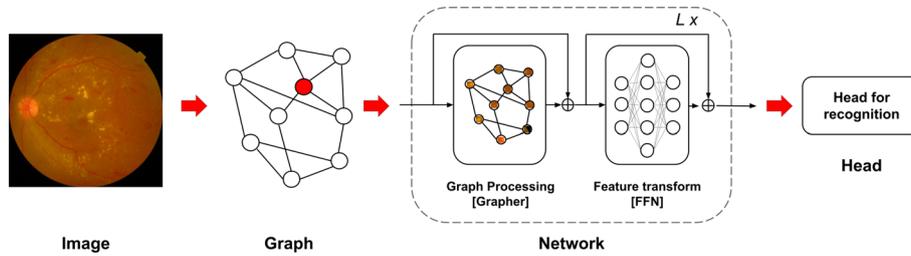**Fig. 9** Augmentations

Fig. 10 ViG model framework

adaptability in a wide range of scenarios. In every epoch, each image was assigned a random selection from the pool of four augmentations or no augmentation at all.

Graph convolutional network: ViG

GCN is a form of convolutional neural network that can operate directly on graphs and exploit their structural properties. GCN may be trained for visual tasks by extracting graph-level information from photos. GCN learns by altering and transferring data across all nodes. This work followed the GCN-based Vision GNN (ViG) model for DR grading which was proposed by Han et al. [53]. Figure 10 shows the overall structure of the ViG model framework. First, images are converted into graph structures by dividing them into different patches. The converted graph representation of the image then enters the ViG network, which consists of two fundamental modules: the Grapher and FFN. The Grapher module contains graph convolution, where the aggregation and update processes occur between graph nodes. The FFN module transforms graph node features, promotes node diversity, and mitigates the over-smoothing issue of standard GNNs.

Figure 11 shows the overall ViG network block used in this work. The ViG network block is a stack of DeepGCN blocks. Each pair of Grapher and FFN modules forms a DeepGCN block. Both the Grapher and FFN modules have skip connections. The Grapher module consists of 2D convolutional layers (conv2d) with 48 kernels of size 1x1, a Graph Convolution layer, and batch normalization layers. Similarly, the FFN module also includes conv2d layers with the same configuration as the Grapher module, along with batch normalization layers and a Gelu activation function. More detailed elaboration of ViG on input data is provided below:

To build a graph structure of an image, dimensions $H \times W \times 3$ get split into N patches. Every patch undergoes a transformation, resulting in a feature vector represented as $x_i \in \mathbb{R}^D$, a set of features $X = [x_1, x_2, \ldots, x_N]$ obtained where i ranges from 1
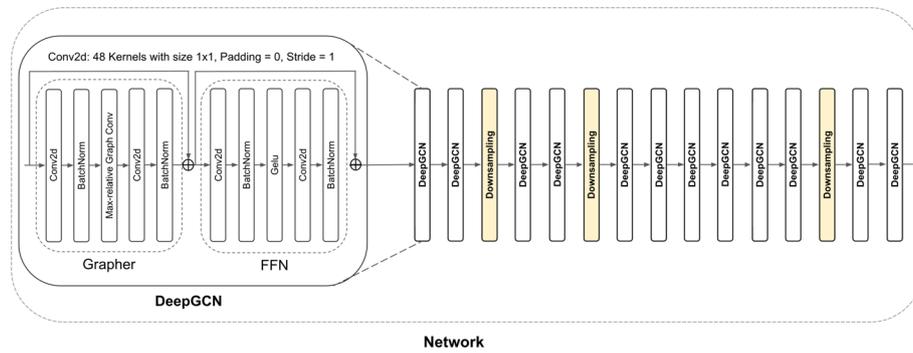
Fig. 11 ViG Network layers

to N and D represents feature dimension. The features are interpreted as a set of unordered collections of nodes $V = \{v_1, v_2, \ldots, v_N\}$. To establish connections between nodes, K-nearest-neighbors $N(v_i)$ get identified for every v_i and create directed edges e_{ji} from v_j to v_i for all $v_j \in N(v_i)$. This process yields a graph $G = (V, E)$, where all edges are represented as E. The above whole process of constructing the graph can be denoted as $\mathcal{G} = G(X)$.

The main Graph-level processing $X' = GraphConv(X)$ was initiated with the feature matrix $X \in \mathbb{R}^{N \times D}$. Using these features the first step involved constructing a graph: $\mathcal{G} = G(X)$. Graph convolutional layer facilitates the transfer of data among nodes by gathering features from neighboring nodes and it functions as follows [53],

$$\mathcal{G}' = F(\mathcal{G}, \mathcal{W}) = \mathcal{U}(\mathcal{A}(\mathcal{G}, W_{\mathcal{A}}), W_{\mathcal{U}}), \tag{1}$$

where, the learnable weights $W_{\mathcal{U}}$ and $W_{\mathcal{A}}$ are used for the *update*($\mathcal{U}$) and *aggregation*($\mathcal{A}$) operations, respectively. $F(\mathcal{G}, \mathcal{W})$ represents a general graph convolution operation at the l-th layer. This can be further derived as $\mathcal{U}(\mathcal{A}(\mathcal{G}, W_{\mathcal{A}}), W_{\mathcal{U}})$ for a more detailed explanation. $\mathcal{G}$ is the input graph and $\mathcal{G}'$ is the output graph at the l-th layer. Aggregation operation $\mathcal{A}$ combines the features of neighboring nodes from graph $\mathcal{G}$ and learnable weights $W_{\mathcal{A}}$ to compute the representation of a node, while update operation $\mathcal{U}$ computes new node representation by applying a non-linear transform on aggregated data to further combines the aggregated feature generated from aggregation operation $\mathcal{A}$ and learnable weights $W_{\mathcal{U}}$. At each layer, the representation of nodes is calculated by aggregating neighbor node features for all nodes as below [53],

$$x_i' = h\big(x_i, g(x_i, N(x_i), W_{\mathcal{A}}), W_{\mathcal{U}}\big), \tag{2}$$

where, $N(x_i)$ is the set of neighbor nodes of x_i at the l-th layer, h is a node feature update function, g is a node feature aggregation function, x_i is node features at l-th layer and x_i' is node features at $l + 1$-th layer. In the node feature aggregation function g, the set of neighbor nodes $N(x_i)$ of x_i got aggregated with learnable weights $W_{\mathcal{A}}$. Later, the aggregated data from aggregation function g got further merged with x_i and learnable weights

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 17 of 37

$W_{\mathcal{U}}$. For efficiency and simplicity, Max-Relative Graph Convolution [66] gets adopted in aggregation process:

$$g(.) = x_i'' = \left[x_i, max\{x_j - x_i | j \in N(x_i)\}\right], \tag{3}$$

$$h(.) = x_i' = x_i'' W_{\mathcal{U}}. \tag{4}$$

In Equation (3) and Equation (4) [53], the bias is ignored. In Equation (3), a max-pooling node feature aggregator is used to pool the differences between the features of node x_i and those of all its neighboring nodes x_j. In Equation (4), the resulting aggregation data x_i'' further went on node feature update function h to generate node features x_i' at $l + 1$-th layer. Besides, in Equation (5), a multi-head update process of graph convolution was introduced. The aggregated feature x_i'' is divided into h number of heads, namely $head^1, head^2, \ldots, head^h$, and each head is updated with its own set of weights $W_{\mathcal{U}}^1, W_{\mathcal{U}}^2, \ldots, W_{\mathcal{U}}^h$. It is possible for every head to update simultaneously and then integrated to form the final value [53]:

$$x_i' = \left[head^1 W_{\mathcal{U}}^1, head^2 W_{\mathcal{U}}^2, \ldots, head^h W_{\mathcal{U}}^h\right] \tag{5}$$

The model may update information in several representation subspaces utilizing the multi-head update procedure, which benefits feature diversity.

The over-smoothing issue in deep GCNs [67, 68] reduces the uniqueness of node features. Therefore degrades visual recognition ability. To address this issue, before and after applying the graph convolution, a linear layer is incorporated to project the node features onto a common domain, promoting feature diversity. After the graph convolution, a nonlinear activation function is introduced to prevent the collapse of layers. This enhanced module is referred to as the Grapher module [53]:

$$Y = \sigma(GraphConv(XW_{in}))W_{out} + X, \tag{6}$$

where, $Y \in \mathbb{R}^{N \times D}$, fully connected layers weights are W_{in} and W_{out}, activation function is σ (GeLU). Bias is ignored. Previously, in Equation (1), $F(\mathcal{G}, \mathcal{W})$ represented a general graph convolution operation at the l-th layer. Equation (2) to Equation (5) were used to further elaborate on this general graph convolution operation at the l-th layer. The whole graph convolution operation for all the layers is represented with $GraphConv(XW_{in})$ function. A non-linear activation function σ is used on the output of $GraphConv(XW_{in})$ function. To tackle the degradation problem, the information from the previous graph convolutional module was transferred to this module through the matrix addition of X.

Feed-forward networks (FFN) are used on each node to boost feature transformation capacity and the over-smoothing problem. The FFN module constitutes a two fully-connected multi-layer perception [53]:

$$Z = \sigma(YW_1)W_2 + Y, \tag{7}$$

where, $Z \in \mathbb{R}^{N \times D}$, fully connected layers weights are W_1 and W_2, and bias is ignored. The output from the Grapher module, Y, was multiplied by the fully connected layer weight W_1 before a non-linear activation function σ was applied. Subsequently, W_2 was

applied, and a skip connection from the previous Grapher module was incorporated into this module through the matrix addition of Y. The hidden dimension of FFN is larger than the feature dimension D. Batch normalization is conducted following each graph convolution and fully-connected layers in both the Grapher and FFN modules, which is removed in Equation (6) and Equation (7) for simplicity. The ViG block is a stack of Grapher and FFN modules. These modules are the fundamental building blocks for developing a Network. The constructed ViG network is illustrated in Fig. 10 according to the above procedures. The Network block in Fig. 10 is further illustrated with details in Fig. 11.

Human activity recognition (HAR)

The GAT classifier is effectively implemented to recognize human actions in this stage.

Sensor data processing

80% of the sensor data were used for training and the rest for testing. There were no null values which eliminated the need for any preparation to handle missing values. Following that, outliers were looked for to find and remove unusual patterns in data arrangement.

$$Upper\ limit = Q3 + 1.5(Q3 - Q1) \tag{8}$$

$$Lower\ limit = Q1 - 1.5(Q3 - Q1) \tag{9}$$

where, $Q1$, $Q2$, $Q3$ are lower, middle and upper quartiles respectively. Data points that lie below the lower limit or above the upper limit are potential outliers. For every feature, the outliers were replaced with the mean of that axis value.

A z-score normalization was used on each data point x to keep the axis values in a standard scale.

$$x' = \frac{(x - \mu)}{\sigma} \tag{10}$$

where, mean is μ, standard deviation is σ and x' is normalized value of data. A new feature was added by calculating the magnitude of three accelerometer axes,

$$magnitude = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{11}$$

where, a_x, a_y and a_z are accelerometer values of x, y and z axes, respectively.

Graph attention network

GAT is a form of Graph Neural Network (GNN) that uses attention processes to learn features from graphs. GAT provides a more complex way of gathering neighborhood information than typical GNNs. This model gives an attention value to each neighbor, signifying how important the neighbor's features are for the node's feature update. These coefficients are calculated employing a shared self-attention method, which assigns an attention score to each pair of nodes. The scores are then standardized across each node's neighborhood with the SoftMax function.
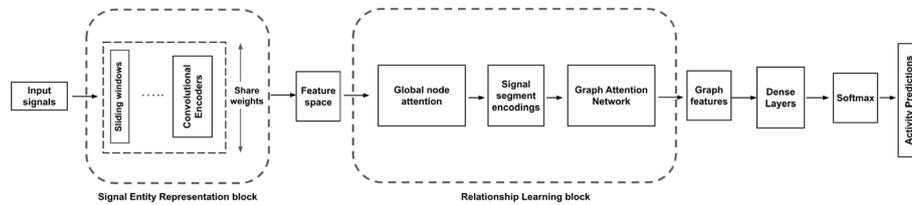
Fig. 12 Graphsensor overall architecture

Among the models explored in this work, the GAT-based Graphsensor from the authors Ge et al. [69] showed the best performance. Figure 12 illustrates the overall architecture of Graphsensor.

The two main components of their proposed technique are relationship learning $\mathcal{G}_{rl}(\mathcal{V}) = \mathcal{V}_{mh}$ and signal segment representation $\mathcal{F}_{sr}(\mathcal{S}) = \mathcal{V}$, where $\mathcal{V}_{mh}$ = multi-head features, and $\mathcal{V}$ = feature map.

Following the methodology from Ge et al. [69], the sensor data $X = \{x_1, x_2, .., x_l\} \in \mathbb{R}^{1 \times L}$ got split by overlapping sliding window with constant length $N (N \in \mathbb{N}, N > 1)$ and overlap rate $P (P \in (0, 1))$ to generate a set of signal segment $\mathcal{S} = \{s_1, s_2, .., s_K\} \in \mathbb{R}^{K \times D}$. Here,

L = Length of the time series
K = Number of signal segments
D = Signal segments dimension

Signal segments are used to evaluate the inner relationships in a particular time series. The length of the signal segment is actually the length of the sliding window N. Equation (12) represents the calculation of the number of signal segments K below [69],

$$K = \left\lfloor \frac{L - N \times P}{N - N \times P} \right\rfloor. \tag{12}$$

Different combinations of sliding window size (N), overlap rate (P), and time series data length per epoch (L) were tested to find the optimal parameter values (shown in Table 4). The accelerometer sensor data length per epoch (L) for each axis was 64 throughout the training process of the HAR dataset. From this, 15,001 signal segments (K) were created using Equation (12) with a window size of 40 datapoints (N) and a 0.99996 overlapping rate (P). Signal segment representation is a convolutional encoder that involves mapping signal segments $\mathcal{S}$ to a feature space $\mathcal{V} = \{v_1, v_2, .., v_K\}$. The convolutional encoder contains three 1D convolution layers, the first of which uses a large kernel with a size of 1x49 to improve the performance of CNN by greatly extending its effective receptive field [70]. The feature map is then extracted from the remaining 2 layers using a smaller kernel with a size of 1x7 after that. A residual Squeeze-and-Excitation block [71] is used at the encoder's end to recalibrate the features discovered by earlier convolution layers.

As internal and external factors impact human behavioral changes over time, global node attention is implemented following the signal segment representation to learn the key factor vector. To learn the key factor vector, the output is then utilized as an input to a two-layer fully connected network.

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 20 of 37

In graph-based self-attention under the Graph Attention Network (GAT) block, the edge between the signal segments is formed by the adjacency matrix, which is produced to depict the connection relationship between the signal segments based on Pearson's correlation coefficient.

The softmax normalized attention method in this work was inspired by [38]. The coefficients a_{ij} calculated by the attention mechanism can be represented as below,

$$a_{ij} = \frac{exp\big(Sigmoid\big(\mathbf{a}\big[pos_1 \times v_i || pos_2 \times v_j\big]\big)\big)}{\sum\limits_{k \in N_i} exp\Big(Sigmoid\Big(\mathbf{a}[pos_1 \times v_i || pos_2 \times v_k]\Big)\Big)}, \tag{13}$$

where, the positional encodings are represented by pos_1 and pos_2, $||$ represents the concatenation operation, a_{ij} is attention coefficient, v_i, v_j and v_k are three different features from the feature space $\mathcal{V}$, and *Sigmoid* is used as nonlinear activation function. Self-attention mechanism $\mathbf{a}$ is a two-layer fully-connected neural network. The dimension of the new signal segment representation matrix is projected using an adaptive average pooling and then utilized as an input to a two-layer fully-connected network to learn the attention coefficient vector. The attention coefficient matrix is created by permuting the attention coefficient vector. The positional encoding used in this work is a development of prior NLP positional encoding techniques [72].

In the case of multi-head attention, a convolutional layer is used to extract important characteristics from the multi-head matrix, which is seen as one feature map. The dense layers and the attention layers together make up the two blocks of the self-attention module. In order to project the signal segments to a higher dimension space, 1x1 convolutions are first performed in the dense layer. A multi-head approach is then used to sustain self-attention learning. In the first run, the multiple-head attention block is stacked many times. To enlarge the input tensor, a dummy axis is employed, changing the input size.

Permuting the signal segment channels from the second dimension to the first dimension allows to extraction of signal segment features from $\mathcal{V}_{mh}$. Then, in order to maintain the identity of the signal segments, feature extraction is carried out using a depthwise convolution with a group size equal to K and padding set to half of the kernel size. The features are subsequently mapped using a two-layer feedforward convolutional neural network. Additionally, a convolutional layer with the same stride and padding parameters as the attention layers are used to map the input dimension to the output. This makes it easier to stack skip connections. H identical layers are stacked to make up the multi-head attention. After the H identical layers, an adaptive average pooling consolidates the feature map of several heads into a single head.

Working with experimental subjects

In this phase, the trained GAT model was applied on 30 days of label-less sensor data collected from four experimental subjects to monitor their activity patterns as part of the evaluation for Diabetes risk factors. Simultaneously, retinal fundus images were gathered for DR assessment using the trained ViG.
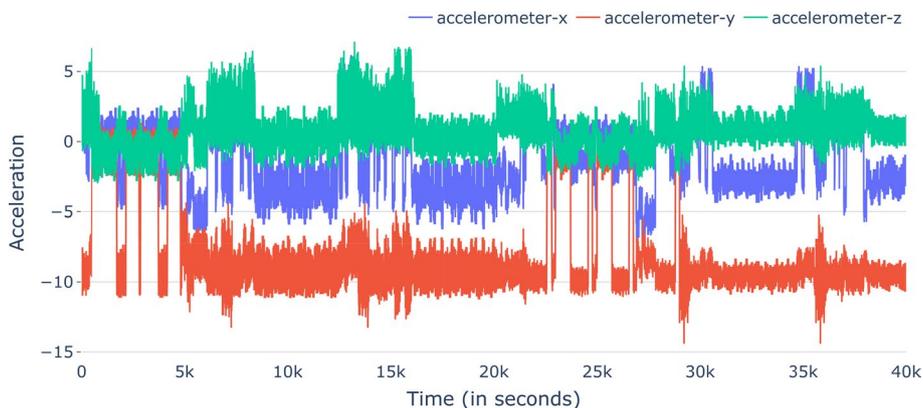
Fig. 13 A continuous eleven hours sample of unlabelled triaxial accelerometer sensor data from experimental subject-1

The Android application was used to gather unlabeled sensor data continuously for 30 days. Figure 13 shows an eleven-hour continuous data sample from subject-1. Initially, height and weight were collected for calculating BMI. Despite potential challenges like constant smartphone attachment to the body of experimental subjects and app running without crashing, smooth data collection was ensured. Data for experimental subjects were recorded at 1 Hz, theoretically amounting to 2,592,000 instances for each subject. However, due to issues like unexpected shutdowns, data totaling 928,182 s, 1,152,713 s, 1,233,233 s, and 951,728 s was obtained from subject-1, subject-2, subject-3, and subject-4, respectively. On average, this amounted to approximately 8.59, 10.67, 11.42, and 8.81 h of data per day for each subject. Despite these obstacles, these data were sufficient for training GAT and predicting the activity patterns of the subjects. Following techniques from Gonzalez [73], high-quality retinal images were captured using a smartphone and a 28D lens.

After applying preprocessing to both types of data collected from experimental subjects, ViG was used for grading DR in retinal images and GAT for predicting activities based on sensor data from the last 30 days. Analyzing occurrences of specific activities allowed us to calculate the time spent on each activity. The sensor data, collected at 1 Hz frequency, provided total occurrences representing seconds spent on activities over the last 30 days (shown in Fig. 14). This data was later converted into minutes. Additionally, the average time interval for each activity per day was determined by dividing the total minutes for each activity over thirty days. This process yielded the daily average duration of each activity. From Fig. 14a, it can be observed that subject-1 spent 90,955 s in cycling (A17) over 30 days. Converting this into minutes equates to $\frac{90,955}{60} = 1515.917$ minutes. That means the subject-1 spent 1515.917 minutes on cycling over 30 days. Hence, the daily average duration for cycling would be $\frac{1515.917}{30} = 50.53$ minutes. Similarly, the daily average duration for each activity for other activities for an experimental subject was calculated. At present, there are DR grade reports alongside the average time intervals of activities in hands from the experimental subjects.
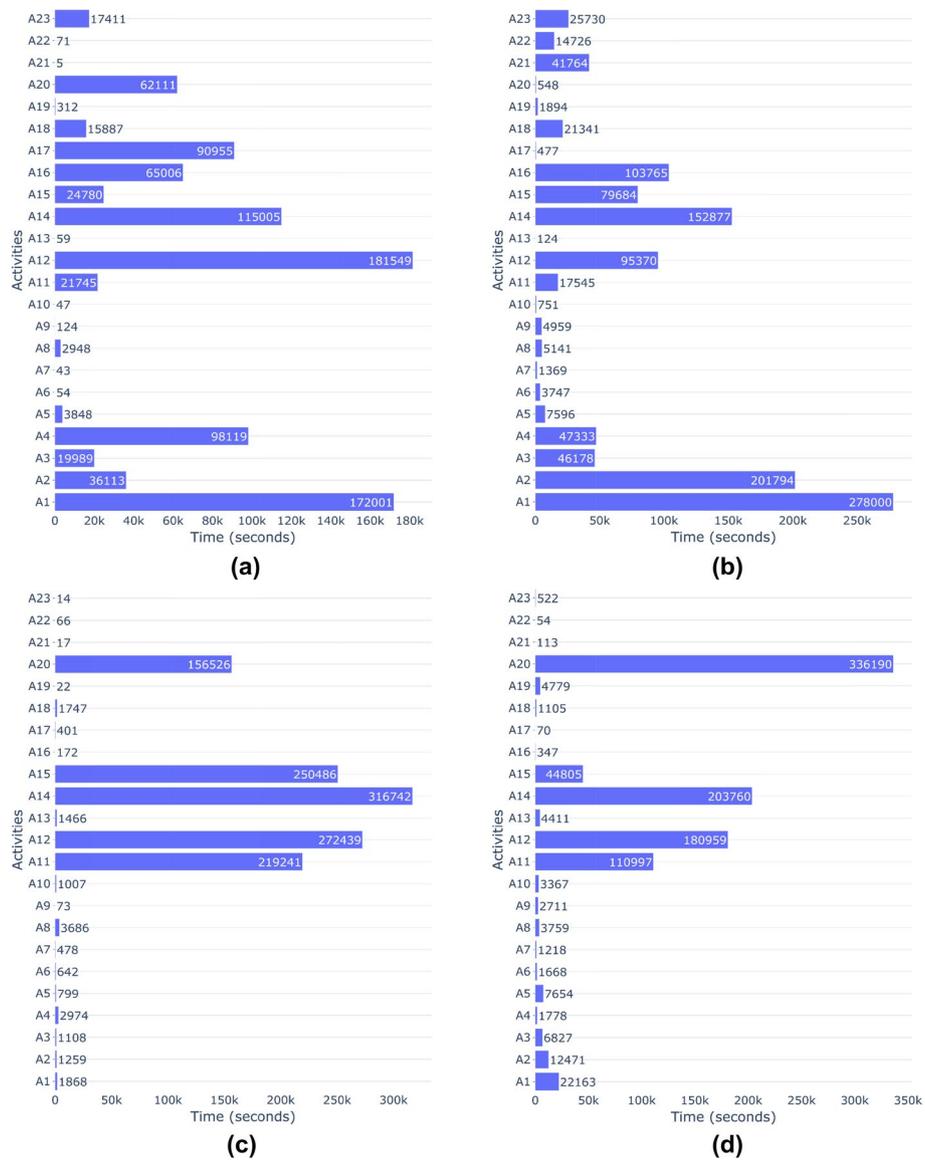
Fig. 14 Total elapsed time (in seconds) of each activity from experimental (**a**) Subject-1, (**b**) Subject-2, (**c**) Subject-3, and (**d**) Subject-4 over 30 days

Similarity measurement and risk factor assesment

In this phase, the Diabetes risk factor was assessed using Cosine similarity by comparing the data from the experimental individuals with the data from diabetic patients. The aim was to comprehend the amount of time diabetic patients devoted to various physical activities before being affected by NIDDM. Data were gathered from 97 diabetic patients at the Sylhet Diabetic Hospital over two months. Individual interviews were conducted. The questionnaire used during the interviews included inquiries about secondary information, such as the patient's gender, weight, height, family history of Diabetes, blood pressure status, and time spent on physical activities. Acknowledging that NIDDM is influenced by genetic factors and lifestyle choices

[74], the secondary information about physical activities was considered primary for this work.

Cosine similarity considers the angle between two objects, taking into account their characteristics as components of the vector. To ensure fair comparisons across dimensions, the measurements were pre-processed using min-max normalization. Each data row of an individual was transformed into its corresponding vector representation. The similarity measurement is defined based on the values of θ. When $\cos\theta = 1$, the two vectors are similar, and when $\cos\theta = 0$, it indicates that the two vectors or objects have no similarity. The Cosine similarity between two objects can be expressed as follows:

$$\cos\theta = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{p} A_i B_i}{\sqrt[2]{\sum_{i=1}^{p} A_i^2} \sqrt[2]{\sum_{i=1}^{p} B_i^2}} \tag{14}$$

Here, A_i and B_i represent segments of vectors A and B, respectively, while A is the experimental subject activity vector and B is the diabetic patient activity vector. $\|A\|$ and $\|B\|$ denote the Euclidean norms of vectors A and B. P is the total number of components, amounting to 28 in the context of this work including age, gender, percentage of Diabetes in first-degree relatives, blood pressure, BMI, and the 23 activities.

When the similarity value ($\cos\theta$) exceeds 75%, the system deduces a likelihood of the subject being significantly affected by Diabetes. Conversely, a similarity value ($\cos\theta$) of 35% indicates a normal case. Therefore, the system established that a similarity ($\cos\theta$) greater than 75% corresponds to a high-risk factor level, while a percentage below 35% corresponds to a low-risk factor level. Moreover, any value of $\cos\theta$ between 35% and 75% is considered a moderate risk factor.

Fig. 15 Early Diagnosis decision matrix. Here, NEDR = no early diagnosis required, EDR = early diagnosis required

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 24 of 37

Fusion for decision

At this point, risk factors and DR grade reports are available for the experimental subjects. A final decision will be reached by fusing these two reports with the help of Fig. 15. As DR is a complication of NIDDM, it can be assumed that someone who has DR also has NIDDM. On that basis, the risk factor and DR grade report were fused to reach a final decision and make the DiabSense system more accurate.

Figure 15 shows the early diagnosis decision matrix where "NEDR" stands for "no early diagnosis required" and "EDR" stands for "early diagnosis required". In the matrix, if the risk factor is "Low" and the DR report shows "No DR" for a subject, the subject will not need an early diagnosis; therefore, "NEDR" will be recommended by DiabSense.

However, for subjects with a "Moderate" or "High" risk factor and a DR report of "No DR", "EDR" will be recommended by DiabSense, indicating that early diagnosis will be required. That means the subject needs medical testing or treatment procedures as soon as possible. This is because the subject might not have the DR complication yet, but they are at a moderate or high risk of developing NIDDM.

If the DR report from the matrix is "Mild DR to PDR". In that case, it will mean that the subject is already experiencing complications from diabetes, regardless of the risk factor being "Low", "Medium", or "High". In such cases, early diagnosis is required. Specifically, if the DR report is "Mild DR to PDR" and the risk factor is either "Low", "Medium", or "High", DiabSense will recommend "EDR", signifying the need for early diagnosis. However, incorporating the risk factors in the decision matrix reveals various categories such as Diabetic with Mild DR condition, Diabetic with Moderate DR condition, Diabetic with Severe DR condition, and Diabetic with Proliferative DR condition. Each category requires medical testing and treatment for Diabetes and DR, tailored to the severity of the condition.

Results and discussion

Classification report on DR grading

Upon fine-tuning the hyperparameters, ViG achieved a classification accuracy of 81.18% for the Aptos 2019 dataset. Fine-tuning was done in the manual searching process where different combinations of hyperparameter values were selected to evaluate the performance of the model during training. The final optimal hyperparameters for the models trained on the Aptos 2019 dataset are shown in Table 3. Figure 16 shows the training session accuracy and loss curve of GCN based ViG model. We have also trained

Table 3 Optimal training hyperparameters for Aptos 2019 dataset

Hyperparameters	ViG	EfficientNet-B5
Optimizer	Adam W	Adam W
Batch size	64	16
Learning rate	$1e-3$	$1e-3$
Weight decay	$5e-2$	$5e-2$
K-neighbor	9	N.A.
Exponential moving avg	0.99996	0.99996
Epochs	114	34
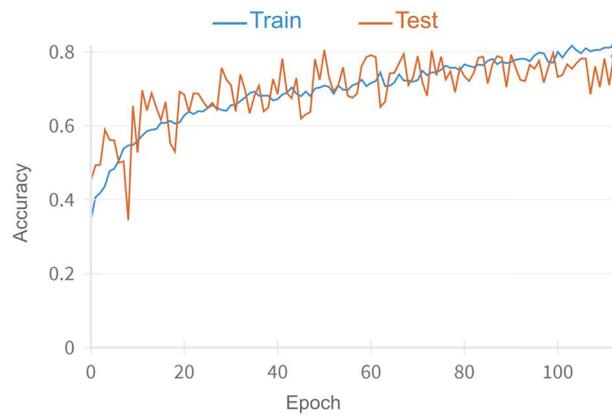
(a)
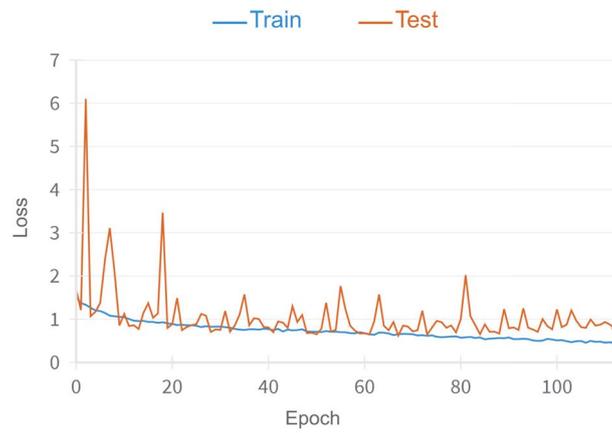
(b)

Fig. 16 **a** Accuracy and **b** Loss progression curves of ViG model in Aptos 2019 data
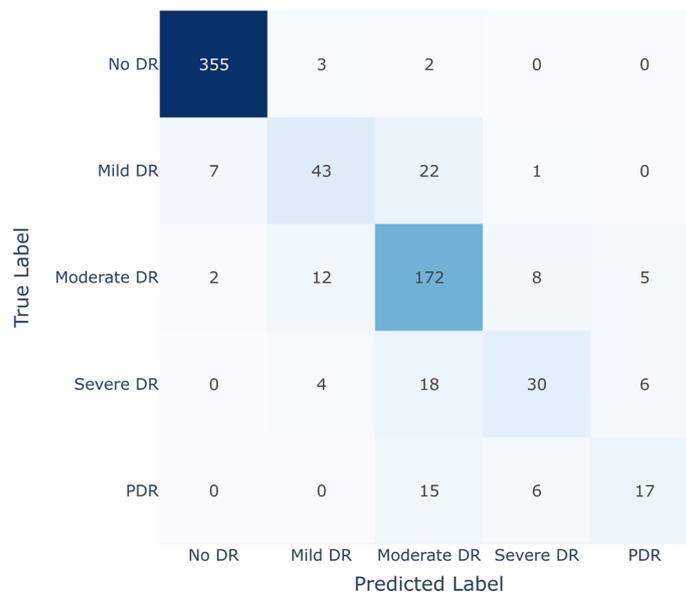
Fig. 17 Confusion matrix of ensembled ViG + EfficientNet-B5

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 26 of 37

a ConvNet-based EfficientNet-B5 and achieved 83.79% for the Aptos data. A soft voting ensemble was applied. As a result, an accuracy of 84.48% was achieved, indicating a notable performance improvement (shown in Table 6).

In medical image analysis, maximizing sensitivity is crucial, as it ensures accurate identification of affected patients. The primary objective of this system is to detect prediabetic and diabetic patients and encourage them to undergo medical testing before their condition worsens. Hence, it is crucial not to misclassify any of the 4 DR conditions as No DR/Healthy which the ensemble of ViG and EfficientNet-B5 has successfully achieved (Fig. 17 confusion matrix validates this). Table 6 demonstrates that ViGs weighted F1-score, precision, and sensitivity are all close enough, while EfficientNet-B5 also maintains consistent metric scores, indicating stable performance for both models. However, their ensemble performance surpasses both individual models. As evident from Table 6, the proposed ensemble outperformed the state-of-the-art models.
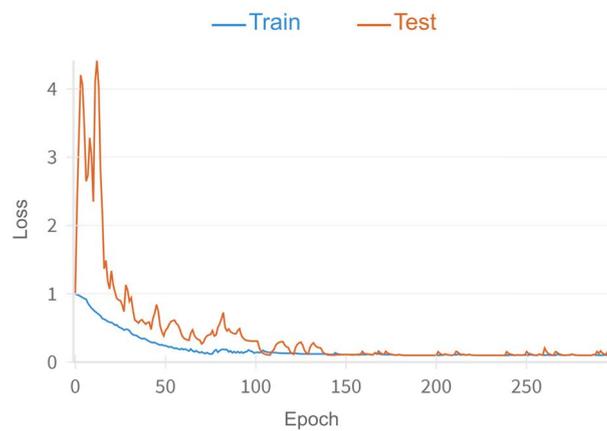
Classification report on HAR

The GAT-based Graphsensor model achieved 98.32% classification accuracy with a f1-score of 97% on the HAR test dataset. Figure 18 shows the accuracy and loss curve and Fig. 20 shows the confusion matrix. The optimal hyperparameters after fine-tuning for the GAT model and other traditional models are listed in Table 4. Similar to the DR phase, fine-tuning involved manually selecting various combinations of hyperparameter values to evaluate the model's performance during training. Since the HAR dataset is newly developed in this work from smartphone triaxial accelerometer sensors with a frequency of 1 Hz, there was no previous work to compare the performance of the GAT model on this dataset. Therefore, to assess the proposed method against conventional approaches, standard DBN, CNN, and RNN-based experiments were conducted as shown in Fig. 19. Clearly, the GAT-based model outperformed the conventional approaches. To check the robustness of this approach, the model was applied to the WISDM dataset as testing data, yielding an accuracy of 96%. This shows that the GAT-based model also outperforms other traditional approaches on the WISDM dataset i.e. 93.32% accuracy Ignatov [31] with CNN, 95.85% accuracy of Xia et al. [35] with LSTM. More comparisons with existing results were shown in Table 7. Thus, the suggested GAT based Graphsensor model exhibited a high percentage of success and robustness in identifying the symptomatic activities of Diabetes.

Table 4 Optimal training hyperparameters for HAR dataset

Hyperparameters	GAT	RNN	LSTM	CNN-LSTM
Optimizer	Adam	Adam	Adam	Adam
Batch size	64	64	128	64
Learning rate	$1e-3$	$1e-3$	$1e-3$	$1e-3$
Weight decay	$1e-4$	$1e-2$	$1e-2$	$1e-2$
Sliding window size (N)	40	N.A.	N.A.	N.A.
Overlap rate (P)	0.99996	N.A.	N.A.	N.A.
Time series data length (L)	64	N.A.	N.A.	N.A.
Epochs	300	450	400	367
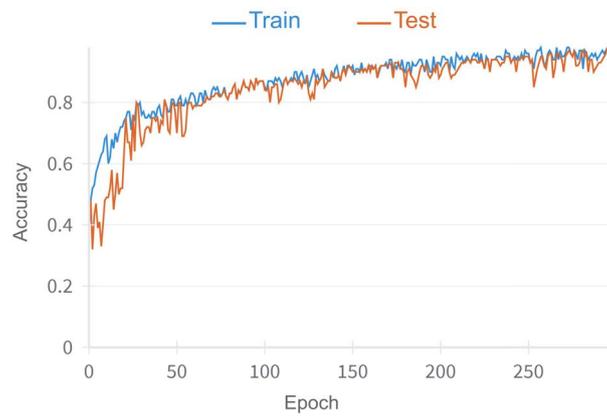
Fig. 18 **a** Accuracy and **b** Loss progression curves for GAT-based Graphsensor model in HAR data
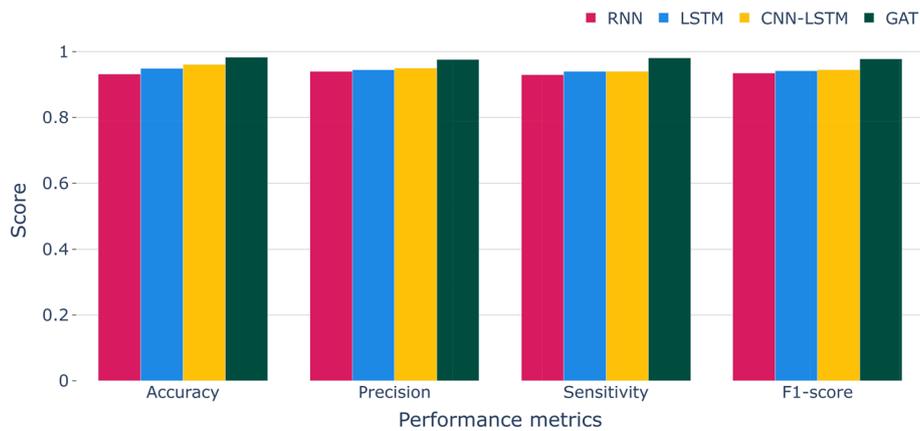

Fig. 19 Classification scores of the GAT model compared with other state-of-the-art models for the HAR dataset of this work (23 classes)

Reports on experimental subjects

As discussed earlier in the methodology, similarity measurement for an experimental subject in this system was performed by calculating the cosine similarity ($\cos \theta$) between
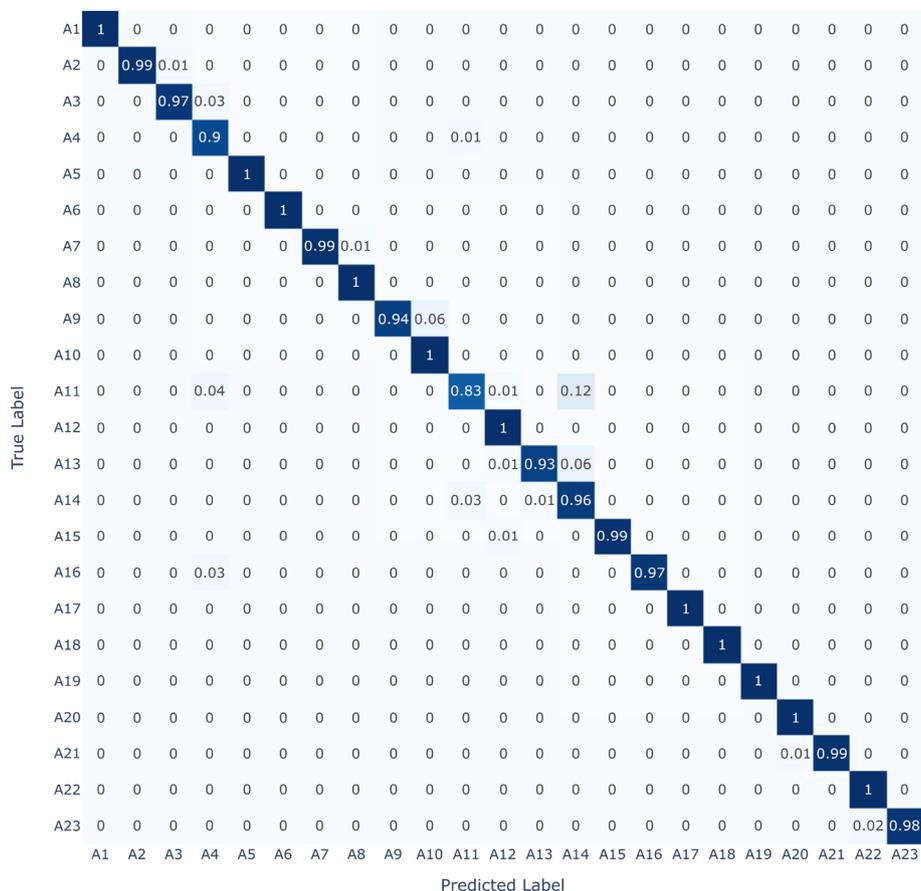
True \ Pred	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
A1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	0.99	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	0.97	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	0.9	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	0	0	0	0	0	0	0.99	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A9	0	0	0	0	0	0	0	0	0.94	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0.04	0	0	0	0	0	0	0.83	0.01	0	0.12	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0.01	0.93	0.06	0	0	0	0	0	0	0	0	0
A14	0	0	0	0	0	0	0	0	0	0	0.03	0	0.01	0.96	0	0	0	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0.99	0	0	0	0	0	0	0	0
A16	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0
A17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
A18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
A21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.99	0	0
A22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
A23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.98

Fig. 20 Confusion matrix of GAT model

the chronicles of average activities of an experimental subject and data from diabetic patients. Furthermore, as discussed in "Similarity measurement and risk factor assessment" section, a similarity value ($\cos\theta$) of 35% or lower between an experimental subject and diabetic patient data indicates a normal case, meaning the subject is not at risk of diabetes. In such cases, the system establishes a risk factor value of "Low". Conversely, when the similarity value ($\cos\theta$) exceeds 75%, the system deduces a likelihood of the subject being significantly affected by diabetes. Therefore, the system establishes that a similarity ($\cos\theta$) greater than 75% corresponds to a "High" risk factor level. Additionally, any value of $\cos\theta$ between 35% and 75% is considered a "Moderate" risk factor.

The average similarity ($\cos\theta$) estimates with diabetic patient data for all four experimental subjects were visualized in Fig. 21. Among the four experimental subjects, the $\cos\theta$ value obtained for the first experimental subject (ES 1) was 23.46%, for the second experimental subject (ES 2) was 28.39%, for the third experimental subject (ES 3) was 75.77%, and for the fourth experimental subject (ES 4) was 57.39%. These results indicate that the first two experimental subjects (ES 1 and ES 2) are classified as "Low" risk factors, the third subject (ES 3) as "High", and the fourth subject (ES 4) as "Moderate".

Table 5 Fused early diagnosis report summary of experimental subjects

Experimental Subject	Reports and Decision			
	DR report	Risk factor	DiabSense decision	A1C level (%)
Subject 1 (ES 1)	No DR	Low	NEDR	5.2
Subject 2 (ES 2)	No DR	Low	NEDR	5.5
Subject 3 (ES 3)	No DR	High	EDR	6.6
Subject 4 (ES 4)	Severe	Moderate	EDR	6.1

ES experimental subject, *NEDR* no early diagnosis required, *EDR* early diagnosis required
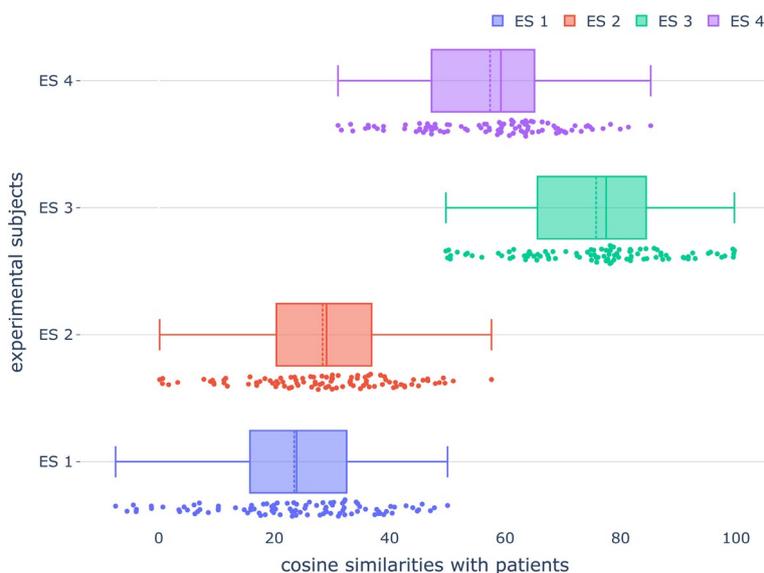
Fig. 21 Correlation between diabetic patients and experimental subjects. *ES* Experimental Subject

Table 6 ViG and ViG+EfficientNet-B5(ensemble) models classification scores compared with previously proposed ANN and ConvNet-based state-of-the-art models for Aptos 2019 data (5 classes)

Method	Accuracy	Precision	Sensitivity	F1-score
VGG16 + CapsNet [64]	75.50	–	–	–
Modified MobileNet [49]	79.01	–	76.47	–
ECOC-SVM [62]	80.70	80.70	80.70	80.70
Blended VGG and Xception + DNN [50]	80.96	–	–	–
Inception-ResNet-v2 [48]	82.18	–	–	–
Composite Gated attention DNN [51]	82.54	82.00	83.00	82.00
Modified Xception [49]	83.09	–	88.24	–
Attention-DenseNet [63]	83.69	–	–	–
Modified VGG16 [84]	84.31	–	–	84.00
LA-NSVM [83]	84.31	75.86	66.16	69.90
ViG	81.18	80.39	81.14	81.00
EfficientNet-B5	83.79	83.09	84.00	83.30
ViG+EfficientNet-B5(ensemble)	84.48	84.07	84.55	84.31

Table 7 Classification scores of the GAT model compared with previously proposed state-of-the-art models for the WIDM dataset

Method	Accuracy	Precision	Sensitivity	F1-score
Bi-LSTM [85]	87.62	87.62	87.62	–
Random Forest [86]	88.14	–	–	–
LMT [87]	90.86	90.00	90.00	90.00
CNN [31]	93.32	–	–	–
BPR Genetic Algorithm [86]	94.02	–	–	–
Genetic Algorithm [88]	95.37	–	–	–
TSE-CNN [89]	95.70	–	–	–
LSTM-CNN [35]	95.85	95.75	95.75	95.78
Bi-LSTM [90]	95.86	–	–	–
CNN-BiLSTM [91]	96.05	–	–	96.04
CNN-GRU [92]	96.41	96.43	96.42	96.39
GAT	96.00	95.93	98.50	97.71

DR classification reports were obtained by feeding retinal fundus images of experimental subjects into the trained ensembled ViG+EffNet model. For the first three experimental subjects (ES 1, ES 2, and ES 3), the classification output was "No DR", indicating they have no diabetic retinopathy. For the final experimental subject (ES 4), the classification output was "Severe", revealing that the subject is suffering from diabetic retinopathy at a "Severe" level.

These experimental subjects' risk factors and DR classification reports were listed in the final report summary (see Table 5). Here, EDR stands for "early diagnosis required", and NEDR stands for "no early diagnosis required". According to the report in Table 5, Subject-1 (ES 1) has a DR report of "No DR" with a "Low" risk factor. Therefore, according to the early diagnosis decision matrix in Fig. 15, this subject did not require an early diagnosis since the subject has a healthy DR condition and a low risk of developing NIDDM. Consequently, "NEDR" is listed in the DiabSense decision column,

Similarly, the DiabSense decision for experimental Subject-2 (ES 2) is "NEDR", as this subject has exactly the same risk factor and DR report as Subject-1 (ES 1).

For Subject-3 (ES 3), the DR report is "No DR", and the risk factor is "High". This means the subject might not have the DR complication yet but is at a high risk of developing NIDDM. Therefore, using the early diagnosis decision matrix in Fig. 15, the DiabSense decision was "EDR".

In the case of Subject-4 (ES 4), the DR report is "Severe", and the risk factor is "Moderate". This indicates that the subject is already suffering from DR and is also at a moderate risk of developing NIDDM. Thus, using the early diagnosis decision matrix in Fig. 15, the DiabSense decision was "EDR".

The results of this system were validated through medical testing, specifically, the A1C test conducted on experimental subjects. A1C offers a dependable indicator of blood sugar levels and has a strong correlation with the likelihood of developing long-term diabetic problems. Pre-analytical and analytical factors are among the many technical benefits that the A1C test offers over the glucose laboratory measures that are currently in use. According to a research [75], A1C testing is standardized and modified to the

DCCT/UKPDS criteria, which makes it different from FPG or 2HPG testing when used to diagnose diabetes. The A1C test has been found to properly reflect continuous glucose levels and has a strong correlation with the risk of complications from diabetes, according to the International Expert Committee for the Diagnosis of Diabetes.

Previous research has suggested that individuals with an A1C level $\leq 5.7\%$ are at no risk of diabetes, A1C levels between 5.7% to 6.4% indicate a pre-diabetic condition. A1C level $\geq 6.5\%$ indicates that the individual has diabetes [75–77]. In Table 5, the DiabSense outcome for Subject-1 and Subject-2 was "NEDR", meaning no early diagnosis is required. The DiabSense outcome for these two subjects was validated through A1C tests. The A1C levels for Subject-1 was 5.2% and for Subject-2 was 5.5%. Since an A1C level $\leq 5.7\%$ indicates no risk of diabetes, the DiabSense outcome successfully corresponds with the A1C medical testing result. The DiabSense outcome for Subject-3 was "EDR", meaning early diagnosis is required. The subject's DiabSense outcome was validated through an A1C test with a level of 6.6%. As an A1C level $\geq 6.5\%$ indicates that the individual has diabetes, the DiabSense outcome is also successfully proportionate with the A1C medical testing result. Similarly, the DiabSense outcome for Subject-4 was "EDR", indicating early diagnosis is required. The subject's DiabSense outcome was validated through an A1C test with a level of 6.1%. Since an A1C level between 5.7% and 6.4% indicates a pre-diabetic condition, pre-diabetic patients need early diagnosis. Hence, the DiabSense outcome successfully fits with the A1C medical testing result. These findings align with the conclusions from the proposed system, indicating a high success rate in early diagnosis recommendations.

Discussion

Studies that used ANN and data mining techniques for early diagnosis of diabetes did not explore the correlation of HAR and DR with NIDDM [78]. The studies focused on diabetic symptomatic datasets collected from patients [78], repository datasets with neuro-fuzzy-based classifiers [79], and patient health records and illness information data [80–82]. In our work, DiabSense demonstrated the correlation between HAR and DR with NIDDM in early diagnosis. Specifically, DiabSense has two components: HAR and DR grading. Prior research worked with traditional ANN-based models such as CNN, LSTM, and their variants. However, DiabSense showed superior performance using GNN-based approaches. Our results indicate that GNN-based approaches in the HAR and DR domains outperformed the traditional ANN-based approaches.

Table 6 shows the comparison of previously proposed state-of-the-art models with our research models for the DR grading phase using Aptos 2019 data. The study proposed in [64] achieved only 75.50% accuracy with a hybrid deep neural network (VGG16 + CapsNet) formed from CNN-based VGG16 and a capsule neural network, despite augmenting the Aptos 2019 dataset to enhance model performance. The CNN-based Modified MobileNet model in [49] showed improved performance with 79.01% accuracy and 76.47% sensitivity. A non-CNN-based approach, the Error Correction Output Code (ECOC)—SVM model proposed in [62], showed a slight performance improvement with 80.70% accuracy and an increased sensitivity score of 80.70%. The Blended VGG and Xception + DNN [50], Inception-ResNet-v2 [48], and Composite Gated attention DNN [51] models showed similar performance in terms of all the performance metrics shown

Alam *et al. Journal of Big Data*   (2024) 11:103

Page 32 of 37

in Table 6. The Modified Xception technique in [49] demonstrated a boost in sensitivity with a score of 88.24%, which is higher than our ensembled model (ViG + EfficientNet-B5), however it falls short in terms of overall accuracy. The Lesion-aware attention with neural support vector machine (LA-NSVM) [83] model showed comparable accuracy to our work but significantly lacked precision, sensitivity, and F1-score. In contrast, our ensemble of ViG and EfficientNet-B5 outperforms all these models with accuracy, precision, sensitivity, and F1-score of 84.48%, 84.07%, 84.55%, and 84.31%, respectively.

Since the HAR dataset is a newly developed dataset from smartphone triaxial accelerometer sensors, there was no previous work for performance comparison. Consequently, the GAT model was applied to the WISDM dataset for a performance benchmark. As shown in Table 7, the Bidirectional LSTM (Bi-LSTM) model by Shan et al. [85] achieved an accuracy, precision, and sensitivity of 87.62%. The Random Forest model proposed by Quaid et al. [86] attained an accuracy of 88.14%, but specific metrics for precision, sensitivity, and F1-score were not provided. The Logistic Model Tree (LMT) by Nematallah et al. [87] achieved an accuracy of 90.86% with a precision and sensitivity of 90.00%. Ignatov [31] achieved an accuracy of 93.32% using a CNN model. Quaid et al. [86] reported an even higher accuracy of 94.02% using a Behavioural Pattern Recognition Genetic Algorithm (BPR Genetic). The Genetic Algorithm model by Jalal et al. [88] achieved a slightly better accuracy of 95.37%, but like the BPR Genetic Algorithm, it did not provide specific metrics for precision, sensitivity, and F1-score. Two different variations of CNN and an LSTM model, TSE-CNN [89], LSTM-CNN [35], and Bi-LSTM [90], achieved very similar accuracies of 95.70%, 95.85%, and 95.86%, respectively. The CNN-BiLSTM [91] model attained an accuracy of 96.05% and an F1-score of 96.04%, which is slightly better in accuracy but lower in F1-score compared to our proposed GAT model. The GAT model, with an accuracy of 96.00%, a precision of 95.93%, a sensitivity of 98.50%, and an F1-score of 97.71%, demonstrates superior performance compared to all previously proposed models. While the CNN-GRU [92] model achieved marginally higher accuracy and precision, the GAT model excels significantly in sensitivity and F1-score, indicating a better overall ability to correctly identify true positives and balance between precision and recall. Therefore, the GAT model is more robust and effective for the WISDM dataset than the previously existing state-of-the-art models.

This study suggests that combining DR and HAR in the early diagnosis of NIDDM holds promising potential for practical clinical practice. The system is smartphone-based, making it accessible to the general population and more cost-effective than regular medical checkups. Since DR is one of many complications of NIDDM, future work could consider additional diabetes-related complications such as nephropathy, neuropathy, heart issues, and oral problems to enhance system robustness, given the clinical relationships between these complications and NIDDM [93].

Conclusion

In this work, diligent efforts were invested in devising a system named DiabSense for early diagnosis of type-2 Diabetes/NIDDM and validated it with medical testing reports. DiabSense shows great promise in effectively managing NIDDM during the early stages. The proposed system integrates human Diabetic Retinopathy (DR) conditions with diabetic risk factors acquired from daily human activity patterns to establish a final decision

for the early NIDDM diagnosis. This reduces the need for expensive routine medical checkups and reliance on symptoms until the system alerts for an early diagnosis positive. As NIDDM symptoms are absent in the early stages for most cases, the undiagnosed patients will get a chance to start their medical diagnosis as soon as DiabSense alerts them. The system relies exclusively on smartphones, eliminating the need for any other expensive components i.e., body sensors or medical fundus cameras. It efficiently consumes power by utilizing data from smartphone sensors at a 1 Hz frequency. The collection of experimental subjects data under varying conditions, phone positions, and intensities to account for energy loss caused by activities, along with the recognition of possible everyday activities related to NIDDM, including irrelevant activities to prevent false recognition, has significantly contributed to the overall robustness of the system. Adept pre-processing techniques and the incorporation of the latest Graph Attention Network (GAT) based Graphsensor and Graph Convolutional Network (GCN) based Vision GNN (ViG) classifiers for HAR and DR recognition tasks respectively further enhanced the systems prediction accuracies from the traditional deep learning approaches. The comparison between DiabSense outcomes and clinical A1C tests confirms its reliability for accurate early diagnosis assessments. According to the results, if the proposed technique is adopted clinically for the early diagnosis of diabetics, millions of human lives can be benefited worldwide.

The proposed system was designed to facilitate the accurate early detection of NIDDM by incorporating both DR and HAR. This objective was successfully achieved, as evidenced by reliable early diagnosis predictions for all four experimental subjects. The models developed for the DR and HAR phases of this system also demonstrated superior accuracy, precision, recall, and F1-score compared to existing models. The proposed ensembled ViG + EfficientNet-B5 model showed improvements of 0.17% in accuracy, 2.07% in precision, 1.55% in sensitivity, and 0.31% in F1-score for DR grading compared to previous models. Additionally, our GAT model achieved a 2.08% improvement in sensitivity and a 1.32% improvement in the F1-score for HAR on the WISDM dataset. These advancements were made possible by the innovative approach of DiabSense. The novel ensembled ViG + EfficientNet-B5 model in DiabSense uniquely considered both regular and irregularly shaped retinal fundus image lesions, unlike existing research that focused solely on regular lesion shapes. Similarly, the GAT model in DiabSense captured the internal connections in time-series tri-axial smartphone sensor data, which previous models neglected by primarily focusing on temporal dependence. DiabSense clearly demonstrates substantial improvements and provides a novel methodology for early NIDDM diagnosis, setting a new benchmark in the field.

One of the few limitations of this work was during the HAR data collection process, where data was only collected when the smartphone was in the participant's pants pocket. However, it is impractical to carry a phone like this constantly. Additionally, individuals might not always have their smartphones with them when they are sleeping. A technique to detect sleep using smartphones must be developed. The Aptos 2019 dataset was developed solely from subjects of Asian descent, hence the HAR dataset was chosen not to include volunteers from other continents, but rather exclusively from Asia due to ethnic factors influencing the risk of NIDDM.

Looking ahead, we aim to broaden the scope of the DiabSense system by considering additional Diabetes-related complications such as nephropathy, neuropathy, heart issues, and oral problems, thereby enhancing its robustness. A multimodal dataset consisting of retinal fundus images and time-series sensor data from the same subjects can be developed and utilized for model training. The HAR dataset can be expanded to include a large number of volunteers from all over the world. Although this research centered on smartphones, integrating handheld devices like smartwatches would further fortify the study.

Declarations

References

1. International Diabetes Federation: IDF Diabetes Atlas, 10th ed. Brussels, Belgium. 2021. https://www.diabetesatlas.org
2. Gregg EW, Li Y, Wang J, Rios Burrows N, Ali MK, Rolka D, Williams DE, Geiss L. Changes in diabetes-related complications in the united states, 1990–2010. N Engl J Med. 2014;370(16):1514–23. https://doi.org/10.1056/NEJMoa1310799.
3. King P, Peacock I, Donnelly R. The uk prospective diabetes study (ukpds): clinical and therapeutic implications for type 2 diabetes. Br J Clin Pharmacol. 1999;48(5):643–8.
4. Simó-Servat O, Hernández C, Simó R. Diabetic retinopathy in the context of patients with diabetes. Ophthalmic Res. 2019;62(4):211–7.
5. Nentwich MM, Ulbig MW. Diabetic retinopathy-ocular complications of diabetes mellitus. World J Diabetes. 2015;6(3):489.
6. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000–2003. J Sci Med Sport. 2004;7(1):6–19.
7. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. N Engl J Med. 1991;325(3):147–52.
8. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, Horton ES, Castorino K, Tate DF. Physical activity/exercise and diabetes: a position statement of the American diabetes association. Diabetes Care. 2016;39(11):2065.
9. Manson JE, Stampfer M, Colditz G, Willett W, Rosner B, Hennekens C, Speizer F, Rimm E, Krolewski A. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. Lancet. 1991;338(8770):774–8.
10. Asia Pacific Tele-Ophthalmology Society: Aptos 2019 blindness detection, 2019. https://www.kaggle.com/c/aptos2019-blindness-detection/data. Accessed 1 Jan 2023
11. Yadav SK, Tiwari K, Pandey HM, Akbar SA. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowl-Based Syst. 2021;223:106970.
12. Han C, Zhang L, Tang Y, Huang W, Min F, He J. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. Expert Syst Appl. 2022;198:116764.
13. Wang Y, Cang S, Yu H. A survey on wearable sensor modality centred human activity recognition in health care. Expert Syst Appl. 2019;137:167–90.
14. Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance. Vis Comput. 2013;29:983–1009.
15. Bahadur EH, Masum AKM, Barua A, Uddin MZ. Active sense: Early staging of non-insulin dependent diabetes mellitus (NIDDM) hinges upon recognizing daily activity pattern. Electronics. 2021. https://doi.org/10.3390/electronics10182194.

16. Barna A, Masum AKM, Hossain ME, Bahadur EH, Alam MS. A study on human activity recognition using gyroscope, accelerometer, temperature and humidity data. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ecce), IEEE, 2019, pp. 1–6.
17. Bodor R, Jackson B, Papanikolopoulos N. Vision-based human tracking and activity recognition. In: Proc. of the 11th Mediterranean Conf. on Control and Automation, Citeseer, 2003;1:1–6.
18. Ni B, Wang G, Moulin P. Rgbd-hudaact: A color-depth video database for human daily activity recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2011;1147–1153.
19. Xia L, Aggarwal J. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013;2834–2841.
20. Aggarwal JK, Xia L. Human activity recognition from 3d data: a review. Pattern Recogn Lett. 2014;48:70–80.
21. Mukhopadhyay SC. Wearable sensors for human activity monitoring: a review. IEEE Sens J. 2014;15(3):1321–30.
22. Nweke HF, Teh YW, Al-Garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst Appl. 2018;105:233–61.
23. Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia, 2015;1307–1310.
24. Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutorials. 2012;15(3):1192–209.
25. Uddin MZ, Soylu A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. Sci Rep. 2021;11(1):16455.
26. Ronao CA, Cho S-B. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst Appl. 2016;59:235–44.
27. Reyes-Ortiz J-L, Oneto L, Samà A, Parra X, Anguita D. Transition-aware human activity recognition using smartphones. Neurocomputing. 2016;171:754–67.
28. Hassan MM, Uddin MZ, Mohamed A, Almogren A. A robust human activity recognition system using smartphone sensors and deep learning. Futur Gener Comput Syst. 2018;81:307–13.
29. Guidoux R, Duclos M, Fleury G, Lacomme P, Lamaudière N, Manenq P-H, Paris L, Ren L, Rousset S. A smartphone-driven methodology for estimating physical activities and energy expenditure in free living conditions. J Biomed Inform. 2014;52:271–8.
30. Ha S, Choi S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE 2016;381–388.
31. Ignatov A. Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl Soft Comput. 2018;62:915–22.
32. Khan A, Hammerla N, Mellor S, Plötz T. Optimising sampling rates for accelerometer-based human activity recognition. Pattern Recogn Lett. 2016;73:33–40.
33. Asim Y, Azam MA, Ehatisham-ul-Haq M, Naeem U, Khalid A. Context-aware human activity recognition (CAHAR) in-the-wild using smartphone accelerometer. IEEE Sens J. 2020;20(8):4361–71.
34. Alsheikh MA, Selim A, Niyato D, Doyle L, Lin S, Tan H-P. Deep activity recognition models with triaxial accelerometers. 2015. arXiv preprint arXiv:1511.04664
35. Xia K, Huang J, Wang H. Lstm-cnn architecture for human activity recognition. IEEE Access. 2020;8:56855–66.
36. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ, et al. Classification accuracies of physical activities using smartphone motion sensors. J Med Internet Res. 2012;14(5):2208.
37. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: International Conference on Pervasive Computing, Springer, 2004;1–17.
38. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, et al. Graph attention networks. Stat. 2017;1050(20):10–48550.
39. Straczkiewicz M, James P, Onnela J-P. A systematic review of smartphone-based human activity recognition methods for health research. NPJ Dig Med. 2021;4(1):148.
40. Umbricht D, Cheng W-Y, Lipsmeier F, Bamdadian A, Lindemann M. Deep learning-based human activity recognition for continuous activity and gesture monitoring for schizophrenia patients with negative symptoms. Front Psych. 2020;11:574375.
41. Islam MM, Nooruddin S, Karray F, Muhammad G. Multi-level feature fusion for multimodal human activity recognition in internet of healthcare things. Info Fusion. 2023;94:17–31.
42. Papadopoulos A, Iakovakis D, Klingelhoefer L, Bostantjopoulou S, Chaudhuri KR, Kyritsis K, Hadjidimitriou S, Charisis V, Hadjileontiadis LJ, Delopoulos A. Unobtrusive detection of Parkinson's disease from multi-modal and in-the-wild sensor data using deep learning techniques. Sci Rep. 2020;10(1):21370.
43. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors-a review of classification techniques. Physiol Meas. 2009;30(4):1.
44. Nathan DM. Long-term complications of diabetes mellitus. N Engl J Med. 1993;328(23):1676–85.
45. Chou C-F, Sherrod CE, Zhang X, Barker LE, Bullard KM, Crews JE, Saaddine JB. Barriers to eye care among people aged 40 years and older with diagnosed diabetes, 2006–2010. Diabetes Care. 2014;37(1):180–8.
46. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.
47. Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, Whitehouse K, Coram M, Corrado G, Ramasamy K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol. 2019;137(9):987–93.
48. Gangwar AK, Ravi V. Diabetic retinopathy detection using transfer learning and deep learning. In: Evolution in computational intelligence: frontiers in intelligent computing: theory and applications (FICTA 2020), 2021;1;679–689. Springer

49. Kassani SH, Kassani PH, Khazaeinezhad R, Wesolowski MJ, Schneider KA, Deters R. Diabetic retinopathy classification using a modified xception architecture. In: 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE 2019;1–6.
50. Bodapati JD, Naralasetti V, Shareef SN, Hakak S, Bilal M, Maddikunta PKR, Jo O. Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. Electronics. 2020;9(6):914.
51. Bodapati JD, Shaik NS, Naralasetti V. Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. J Ambient Intell Humaniz Comput. 2021;12(10):9825–39.
52. Adem K. Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks. Expert Syst Appl. 2018;114:289–95.
53. Han K, Wang Y, Guo J, Tang Y, Wu E. Vision GNN: an image is worth graph of nodes. Adv Neural Inf Process Syst. 2022;35:8291–303.
54. Lee J, Kim J et al. Energy-efficient real-time human activity recognition on smart mobile devices. Mobile Information Systems 2016;2016.
55. Li C, Niu D, Jiang B, Zuo X, Yang J. Meta-har: Federated representation learning for human activity recognition. In: Proceedings of the Web Conference 2021, 2021;912–922.
56. Soni V, Yadav H, Semwal VB, Roy B, Choubey DK, Mallick DK. A novel smartphone-based human activity recognition using deep learning in health care. In: Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021, Springer 2023;493–503.
57. Bhattacharya D, Sharma D, Kim W, Ijaz MF, Singh PK. ENSEM-HAR: an ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. Biosensors. 2022;12(6):393.
58. D'Angelo G, Palmieri F. Enhancing COVID-19 tracking apps with human activity recognition using a deep convolutional neural network and har-images. Neural Comput Appl. 2023;35(19):13861–77.
59. Phukan N, Mohine S, Mondal A, Manikandan MS, Pachori RB. Convolutional neural network-based human activity recognition for edge fitness and context-aware health monitoring devices. IEEE Sens J. 2022;22(22):21816–26.
60. Chakravarthy SS, Bharanidharan N, Kumar VV, Mahesh T, Khan SB, Almusharraf A, Albalawi E. Intelligent recognition of multimodal human activities for personal healthcare. IEEE Access 2024.
61. Nasir D, Bourkha MEA, Hatim A, Elbeid S, Ez-ziymy S, Zahid K. Predicting blood glucose levels in type 1 diabetes using lstm. In: Modern Artificial Intelligence and Data Science: Tools, Techniques and Systems. Springer, Cham Switzerland 2023;121–135.
62. Wong W, Juwono FH, Capriono C. Diabetic retinopathy detection and grading: A transfer learning approach using simultaneous parameter optimization and feature-weighted ecoc ensemble. IEEE Access 2023
63. Dinpajhouh M, Seyyedsalehi SA. Automated detecting and severity grading of diabetic retinopathy using transfer learning and attention mechanism. Neural Comput Appl. 2023;35(33):23959–71.
64. Kumar G, Chatterjee S, Chattopadhyay C. Dristi: a hybrid deep neural network for diabetic retinopathy diagnosis. SIViP. 2021;15(8):1679–86.
65. Google: Case Study: TensorFlow in Medicine—Retinal Imaging, TensorFlow Dev Summit 2017. 2017. https://youtu.be/oOeZ7IgEN4o?t=156
66. Li G, Muller M, Thabet A, Ghanem B. Deepgcns: Can gcns go as deep as CNNS? In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019;9267–9276.
67. Li Q, Han Z, Wu X-M. Deeper insights into graph convolutional networks for semi-supervised learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018;32.
68. Oono K, Suzuki T. Graph Neural Networks exponentially lose expressive power for node classification. arXiv preprint arXiv:1905.10947 2019.
69. Ge J, Xu G, Lu J, Xu X, Meng X. Graphsensor: a graph attention network for time-series sensor. Electronics. 2024;13(12):2290.
70. Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022;11963–11975.
71. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018;7132–7141.
72. Wang B, Zhao D, Lioma C, Li Q, Zhang P, Simonsen JG. Encoding word order in complex embeddings. 2019. arXiv preprint arXiv:1912.12333.
73. González PA. American Academy of Ophthalmology: How to Take Retinal Images with a Smartphone. 2020. https://www.aao.org/education/clinical-video/how-to-take-retinal-images-with-smartphone#disqus_thread].
74. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. Endocr Rev. 2016;37(3):278–316.
75. Committee TIE. International expert committee report on the role of the a1c assay in the diagnosis of diabetes. Diabetes Care. 2009;32(7):1327.
76. Lorenzo C, Wagenknecht LE, Hanley AJ, Rewers MJ, Karter AJ, Haffner SM. A1c between 5.7 and 6.4% as a marker for identifying pre-diabetes, insulin sensitivity and secretion, and cardiovascular risk factors: the insulin resistance atherosclerosis study (iras). Diabetes Care. 2010;33(9):2104–9.
77. Zhang X, Gregg EW, Williamson DF, Barker LE, Thomas W, Bullard KM, Imperatore G, Williams DE, Albright AL. A1c level and future risk of diabetes: a systematic review. Diabetes Care. 2010;33(7):1665–73.
78. Islam M, Ferdousi R, Rahman S, Bushra HY. Likelihood prediction of diabetes at early stage using data mining techniques. In: Computer Vision and Machine Intelligence in Medical Image Analysis, Springer, Cham Switzerland 2020;113–125.
79. Kaur C, Al Ansari MS, Dwivedi VK, Suganthi D. Implementation of a neuro-fuzzy-based classifier for the detection of types 1 and 2 diabetes. Advances in Fuzzy-Based Internet of Medical Things (IoMT), 2024;163–178.
80. Sivaraman M, Thyagarajan M, Sumitha J. Predicting early stage disease diagnosis using machine learning algorithms. In: 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), IEEE 2023;1177–1183.
81. Li X, Zhang J, Safara F. Improving the accuracy of diabetes diagnosis applications through a hybrid feature selection algorithm. Neural Process Lett. 2023;55(1):153–69.

82. Alex SA, Nayahi JJV, Shine H, Gopirekha V. Deep convolutional neural network for diabetes mellitus prediction. Neural Comput Appl. 2022;34(2):1319–27.
83. Shaik NS, Cherukuri TK. Lesion-aware attention with neural support vector machine for retinopathy diagnosis. Mach Vis Appl. 2021;32(6):126.
84. Bodapati JD, Shaik NS, Naralasetti V. Deep convolution feature aggregation: an application to diabetic retinopathy severity level prediction. SIViP. 2021;15:923–30.
85. Shan CY, Han PY, Yin OS. Deep analysis for smartphone-based human activity recognition. In: 2020 8th International Conference on Information and Communication Technology (ICoICT), IEEE 2020;1–5.
86. Quaid MAK, Jalal A. Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. Multimedia Tools Appl. 2020;79(9):6061–83.
87. Nematallah H, Rajan S, Cretu A-M. Logistic model tree for human activity recognition using smartphone-based inertial sensors. In: 2019 IEEE SENSORS, IEEE 2019;1–4.
88. Jalal A, Quaid MAK, Tahir S, Kim K. A study of accelerometer and gyroscope measurements in physical life-log activities detection systems. Sensors. 2020;20(22):6670.
89. Huang J, Lin S, Wang N, Dai G, Xie Y, Zhou J. TSE-CNN: a two-stage end-to-end CNN for human activity recognition. IEEE J Biomed Health Inform. 2020;24(1):292–9. https://doi.org/10.1109/JBHI.2019.2909688.
90. Yin X, Liu Z, Liu D, Ren X. A novel CNN-based BI-LSTM parallel model with attention mechanism for human activity recognition with noisy data. Sci Rep. 2022;12(1):7878.
91. Challa SK, Kumar A, Semwal VB. A multibranch CNN-BILSTM model for human activity recognition using wearable sensor data. Vis Comput. 2022;38(12):4095–109.
92. Lu L, Zhang C, Cao K, Deng T, Yang Q. A multichannel CNN-GRU model for human activity recognition. IEEE Access. 2022;10:66797–810.
93. Kulkarni A, Thool AR, Daigavane S. Understanding the clinical relationship between diabetic retinopathy, nephropathy, and neuropathy: a comprehensive review. Cureus 2024;16(3).

Publisher's Note