IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# GRAPH NEURAL NETWORK-BASED CARDIOVASCULAR RISK PREDICTION USING ELECTRONIC HEALTH RECORD DATA

Mr. P Ankamarao, Dr. G Chamundeswari, Mr. G Hari Hara Kumar, Mrs. B Rajeswari, Mr. B Venkateswara Rao,
Mrs. K Lehamma Department of Computer Science and Engineering
Ramachandra College of Engineering, Eluru

Doi: https://doi.org/10.56975/ijcrt.v14i7.309999

**Abstract** The cardiovascular diseases (CVDs) can be listed among the significant causes of mortality in the world and this is why there is a need of a reliable and early answer to the risk. Since Electronic Health Records (EHRs) are integrated in numerous hospitals, the amount of structured patient data to analyze and predict is enormous. However, the historical machine learning techniques typically assume the existence of individual samples of patient records and ignore the already existing interrelations among individuals within the same clinical phenotype. This is a weakness because it narrows their ability to adopt complex relationships, which are present in real medical records.

In this paper, the proposal of a model based on Graph Neural Network (GNN) and predicting cardiovascular risk alongside the structured EHR data is provided. The proposed model is a patient is represented as a node of a graph and the graph is formed based on similarity data of both clinical and demographic variables that comprise age, blood pressure, cholesterol levels and heart rate. Such representation using graphs helps the model to learn the characteristics of the patients, and the relation that may be possible between the patients.

The architecture uses two modern GNNs, i.e. Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) to train meaningful representations of the constructed graph of patient similarities. The publicly available datasets, including UCI Heart Disease, are tested and trained on the models. Evaluation of the performance utilizes the assistance of such common measurement criteria as accuracy, precision, recall, F1-score, and ROC-AUC.

The experimental findings show that the suggested GNN-based method provides better predictive results than the conventional machine learning models due to the effective use of the relational information in the data. The created system will be able to help medical staff detect high-risk people early enough and facilitate informed clinical judgment. This work indicates how graph-based deep learning methods may be effective to increase predictive analytics and emerging individualized healthcare solutions.

Keyw.ords— Gra.ph Neural Netw.ork (GNN), Cardiov.ascular Disease Predi.ction, Elect.ronic Hea.lth Reco.rds (EHR), Gra.ph Convolu.tional Netw.ork (GCN), Gra.ph Atten.tion Netw.ork (GAT), Deep Lear.ning, Risk Asses.sment, Clin.ical Deci.sion Supp.ort Sys.tem, Machine Lear.ning, Healt.hcare Analy.t

## I Introduction

Cardiovascular dise.ases (CVDs) are com.mon cau.ses of morta.lity in the wor.ld and they have been rais.ing the morta.lity rate per ann.um. Tim.ely preve.ntion and tim.ely provi.sion of medi.cal care to high.-risk indivi.duals with likel.ihood of devel.oping hea.rt rela.ted dise.ases requ.ires ear.ly identif.ication of such indivi.duals. Sin.ce the sph.ere of healt.hcare techno.logies evol.ves at a rap.id pace, Elect.ronic Hea.lth Reco.rds (EHRs) curre.ntly repre.sent a profi.table sou.rce of organ.ised pati.ent data, demogr.aphic information, clin.ical measurements, labor.atory find.ings, and medi.cal hist.ory. This acc.ess to massive healt.hcare data has also pos.ed new vis.tas of crea.ting predi.ctive mod.els to enha.nce clin.ical decisio.n-making.

Conven.tional mach.ine lear.ning meth.ods, inclu.ding, but not limi.ted to, logi.stic regre.ssion and deci.sion tre.es, supp.ort vect.ors mach.ine have been univer.sally appl.ied to pred.ict cardiov.ascular ris.ks. The.se mod.els can det.ect tre.nds in struc.tured data data.sets, and can have suffi.cient predi.ctive pow.er. Noneth.eless, a signif.icant draw.back of such meth.ods is that they tre.at indiv.idual pati.ent reco.rds as sin.gle poi.nts of data. Whe.re it com.es to the act.ual clinical situa.tion, pati.ents usua.lly have simila.rities in their symp.toms, medi.cal hist.ory, and lifes.tyle fact.ors. These relatio.nships can be igno.red, ther.eby los.ing use.ful conte.xtual inform.ation and can low.er the effecti.veness of predi.ction mod.els as a who.le.

The lat.est improv.ements in the fie.ld of deep lear.ning have brou.ght more adva.nced appro.aches to the proce.ssing of comp.lex data. Even tho.ugh neu.ral networ.k-based mod.els are able to represent nonlinear associations among features, they use indepe.ndent samp.les as a prim.ary assumption when used with tabu.lar healthcare data. Gra.ph Neu.ral Netw.orks (GNNs) have ari.sen as an irresi.stible met.hod of the mode.ling of relat.ional inform.ation to over.come this draw.back. Und.er a graph-based mod.el, enti.ties like pati.ents are consi.dered as nod.es, and the.ir corresp.ondence is tak.en as edg.es. This frame.work enab.les the mod.el to acqu.ire pers.onal feat.ures as well as intera.ction betw.een rela.ted enti.ties.

A gra.ph neu.ral netw.ork frame.work is also sugge.sted in this work offe.ring predi.cting cardiov.ascular risk with the help of struc.tured EHR data. The met.hod crea.tes a simil.arity gra.ph of pati.ents by link.ing peo.ple with clin.ical and demogr.aphic characteristics. Sophist.icated GNN mod.els such as Gra.ph Convolu.tional Netw.orks (GCN) and Gra.ph Atten.tion Netw.orks (GAT) are used to lea.rn the meani.ngful gra.ph struc.ture represen.tations. The prop.osed sys.tem will ser.ve to enha.nce predi.ction accu.racy as comp.ared to tradit.ional meth.ods by inclu.ding relat.ional inform.ation.

Some of the contrib.utions in this pap.er are: the des.ign of a graph.-based lear.ning mod.el to pred.ict cardiov.ascular ris.ks, the applic.ation of GCN and GAT mod.els to desc.ribe the relati.onship betw.een pati.ents, and a compar.ative analysis betw.een the traditional mach.ine lear.ning meth.ods and the new appr.oach. The giv.en sys.tem will ass.ist healt.hcare profess.ionals by suppl.ying them with credible data to facilitate ear.ly diagno.stics and indiv.idual treat.ment pla.ns.

## II. RELATED WORK

The top.ic of the mach.ine lear.ning and deep learnin.g-based predi.ction of cardiov.ascular dise.ase has been thoro.ughly resea.rched dur.ing the past yea.rs. Init.ial rese.arch was mos.tly bas.ed on statis.tical techn.iques and traditional mach.ine lear.ning algorithms inclu.ding logi.stic regre.ssion, deci.sion trees, nai.ve Bay.es, support vec.tor mach.ines and so on. The.se meth.ods perfo.rmed reaso.nably in struc.tured data.sets, but they do not have the abil.ity to enc.ode comp.lex nonli.near relat.ions betw.een clin.ical characte.ristics [1] - [4].

In ord.er to overcome the.se constr.aints, scholars have prop.osed ensembl.e-based learning meth.ods like Random For.est and Grad.ient Boos.ting that refi.ned the.ir predi.ction accu.racy by integr.ating nume.rous weak lear.ners. The.se mod.els genera.lized hig.her than the sin.gle classi.fiers and fou.nd exten.sive use in healt.hcare predi.ction syst.ems [5]–[7]. Alth.ough this has been adva.nced, the conven.tional mach.ine lear.ning meth.ods cont.inue to cons.ider pati.ent reco.rds in isola.tion and they do not prov.ide inter-.patient relat.ions.

The conc.ept of usi.ng neu.ral netw.ork mod.els in cardiov.ascular risk predi.ction was fir.st expl.ored when deep lear.ning was devel.oped and used in this fie.ld. Artif.icial Neu.ral Netw.orks (ANNs) have been repo.rted to perf.orm bet.ter when it com.es to non Linear intera.ction of feat.ures wher.eas Convolu.tional Neu.ral Netw.orks (CNNs) and Recur.rent Neu.ral Netw.orks (RNNs) have been invest.igated into captu.ring high-dim.ensional and temp.oral healt.hcare data

[8]–[10]. Even tho.ugh the.se mod.els repre.sent feat.ures bet.ter, they usua.lly ass.ume indepe.ndent samp.les and do not expli.citly deal with the inform.ation that repre.sents relatio.nships betw.een pati.ents.

Graph Neural Networks (GNNs) present an exceptionall as predicting dise.ases, analy.zing simila.rities of pati.ents, and evalu.ating the.ir clin.ical risk [11]–[13]. With the help of gra.ph represe.ntation, the.se mod.els can be able to capture hid.den depend.encies that may.be mis.sed in most tradit.ional meth.ods. Stud.ies on GNN as clinical predi.ctors have demonstrated encour.aging find.ings on Elect.ronic Hea.lth Reco.rds (EHRs). Demogr.aphic and clin.ical characte.ristics have been used to bui.ld pati.ent simil.arity graphs whi.ch have enab.led model constr.uction to deter.mine patt.erns amon.gst pati.ents who share similar medi.cal condi.tions [14]-[16]. Gra.ph bas.ed archite.ctures like the Gra.ph Convolu.tional Netw.orks (GCN) and Graph Atten.tion Netw.orks (GAT) have furt.her enha.nced the perfor.mance of the predictions proc.ess sin.ce they facil.itate effec.tive propag.ation of inform.ation and dynami.cally wei.ght nod.es using the.ir neighbors.

More.over, it has been sugge.sted to use hyb.rid mod.els that comb.ine GNNs with oth.er deep lear.ning mod.els, inclu.ding the Long Short.-Term Mem.ory (LSTM) netw.ork and atten.tion mecha.nisms to capt.ure both spat.ial and temporal depend.encies in healt.hcare data [17]–[18]. The.se meth.ods have sho.wn a hig.her lev.el of accu.racy in long.-term hea.lth risks foreca.sting and chro.nic illnesses.

Alth.ough the.re is incre.asing atten.tion to the idea of graph.-based lear.ning, the sco.pe of liter.ature on the use of GNN mod.els to pred.ict cardiov.ascular ris.ks with the help of struc.tured EHR data is very sma.ll. A signif.icant num.ber of stud.ies curre.ntly perfo.rmed tend to be in general dise.ase predi.ction or unstru.ctured like medi.cal imag.ing. Conseq.uently, a sophist.icated frame.work with the poten.tial to util.ize pati.ent similarity gra.ph and sophist.icated GNN archite.ctures to enha.nce cardiov.ascular dise.ase foreca.sting is requ.ired.

The pap.er is an exten.sion of the exis.ting liter.ature by sugge.sting a Gra.ph Neu.ral Networ.k-based frame.work of the relati.onship betw.een pati.ents bas.ed on struc.tured EHR data. The prop.osed solu.tion will impr.ove the predi.ctive accu.racy and off.er more predi.ctive decision.-support inform.ation in cardiov.ascular serv.ices by combi.ning the capabi.lities of graph.-based lear.ning and clin.ical data.

## III. RESEARCH GAP

Desp.ite signif.icant advanc.ements in mach.ine lear.ning and deep lear.ning techn.iques for cardiovascular dise.ase predi.ction, seve.ral crit.ical limita.tions rem.ain unres.olved in existing appro.aches.

Most tradit.ional mach.ine lear.ning mod.els tre.at pati.ent reco.rds as indepe.ndent data poi.nts and rely sol.ely on featur.e-based representations.

This assum.ption igno.res the inhe.rent effec.tive tool in the past few yea.rs due to the.ir abil.ity to pred.ict relat.ional data. GNNs cons.ider enti.ties as nod.es and relatio.nships as edg.es, whi.ch makes both struc.tural and featur.e-level inform.ation lear.ner. A num.ber of wor.ks have relatio.nships amo.ng pati.ents who may sha.re simi.lar clinical

characte.ristics, lifes.tyle patt.erns, or medi.cal histo.ries. As a res.ult, the.se mod.els oft.en fail to capt.ure hidden depend.encies within healthcare data, lead.ing to subop.timal predi.ction performance in real-world scenarios. sho.wn that GNNs can be effect.ively used in healt.hcare, suchAlth.ough deep lear.ning mod.els such as Artif.icial Neu.ral Networks (ANNs), Convolu.tional Neu.ral Netw.orks (CNNs), and Recur.rent Neu.ral Netw.orks (RNNs) have impr.oved the abil.ity to mod.el nonli.near relatio.nships amo.ng feat.ures, they sti.ll prima.rily oper.ate on isol.ated samp.les when appl.ied to struc.tured EHR data.sets. These mod.els do not expli.citly incorp.orate inter-.patient relatio.nships, whi.ch are cruc.ial for underst.anding dise.ase progre.ssion and risk patt.erns.

Rec.ent rese.arch on Gra.ph Neu.ral Netw.orks (GNNs) has shown promi.sing resu.lts in mode.ling relat.ional data. Howe.ver, the applic.ation of GNN-based appro.aches for cardiov.ascular risk predi.ction rema.ins limi.ted. Exis.ting stud.ies oft.en foc.us on gene.ral dise.ase predi.ction tas.ks or util.ize unstru.ctured data such as medi.cal ima.ges, rat.her than structured Elect.ronic Hea.lth Rec.ord (EHR) data.

Furthe.rmore, many exis.ting predi.ction syst.ems do not adequ.ately addr.ess chall.enges such as data imbalance, feat.ure depen.dency, and lack of interpre.tability. The abse.nce of patient similari.ty-based mode.ling and insuff.icient compar.ative anal.ysis with tradit.ional meth.ods highl.ight the need for a more compreh.ensive and rob.ust frame.work.

Therefore, the.re is a cle.ar rese.arch gap in devel.oping a graph-based deep lear.ning appr.oach that effect.ively capt.ures relatio.nships amo.ng pati.ents using struc.tured EHR data, whi.le also impro.ving predi.ction accu.racy and provi.ding reli.able insi.ghts for clin.ical decisio.n-making.

## IV. OBJECTIVES

The prim.ary objec.tive of this rese.arch is to deve.lop an effic.ient and reli.able cardiov.ascular risk predi.ction sys.tem usi.ng Graph Neural Netw.ork (GNN.)-based techn.iques appl.ied to struc.tured Elect.ronic Hea.lth Rec.ord (EHR) data. The stu.dy aims to impr.ove predi.ction accu.racy by incorpo.rating relat.ional inform.ation amo.ng pati.ents.

The spec.ific objec.tives of this work are as foll.ows:

- To des.ign and imple.ment a graph-based frame.work that repre.sents pati.ents as nod.es and the.ir clin.ical simila.rities as edg.es usi.ng EHR data.
- To prepr.ocess and trans.form struc.tured healt.hcare data thro.ugh handling miss.ing val.ues, encoding catego.rical varia.bles, and normal.izing numer.ical feat.ures to ens.ure data consis.tency and mod.el effic.iency.
- To const.ruct a pati.ent similarity graph usi.ng approp.riate simil.arity meas.ures such as cos.ine simil.arity or Eucli.dean dist.ance bas.ed on clin.ical attri.butes.
- To deve.lop and tra.in adva.nced Gra.ph Neu.ral Netw.ork mod.els, inclu.ding Gra.ph Convolu.tional Netw.ork (GCN) and Gra.ph Atten.tion Netw.ork (GAT), for cardiovascular risk predi.ction.
- To eval.uate the perfor.mance of the prop.osed mod.els usi.ng standard metr.ics such as accu.racy, preci.sion, rec.all, F1-s.core, and ROC.-AUC.
- To comp.are the perfor.mance of graph.-based mod.els with tradit.ional mach.ine lear.ning algor.ithms such as Logi.stic Regre.ssion, Ran.dom For.est, and Supp.ort Vec.tor Mach.ine.

To anal.yze the effecti.veness of graph.-based lear.ning in captu.ring hid.den relatio.nships amo.ng pati.ents and impro.ving predi.ctive perfor.mance in healt.hcare applic.ations.

## V. METHODOLOGY

The prop.osed sys.tem ado.pts a graph.-based deep lear.ning appr.oach for cardiov.ascular risk predi.ction usi.ng struc.tured Electronic Hea.lth Rec.ord (EHR) data. The method.ology cons.ists of mult.iple sta.ges, inclu.ding data colle.ction, preproc.essing, gra.ph constr.uction, model develo.pment, and perfor.mance evalu.ation. The over.all objec.tive is to leve.rage relat.ional lear.ning to impr.ove prediction accu.racy comp.ared to conven.tional appro.aches.

**A.** Dataset Description

The stu.dy util.izes publ.icly avail.able struc.tured healt.hcare data.sets for cardiov.ascular risk predi.ction. The prim.ary data.set consi.dered is the UCI Hea.rt Dise.ase data.set, whi.ch cont.ains clin.ical attri.butes such as age, sex, che.st pain type, rest.ing blood pressure, choles.terol lev.el, fast.ing blo.od sug.ar, electrocardiographic resu.lts, and hea.rt rate. Additi.onally, the Frami.ngham data.set may be used to enha.nce mod.el valid.ation and general.ization.

The.se data.sets prov.ide suffi.cient clin.ical and demogr.aphic inform.ation requ.ired to mod.el pati.ent characte.ristics and const.ruct simil.arity relatio.nships for graph.-based lear.ning.

**B.**Data Preprocessing

Data preproc.essing is an essen.tial step to ens.ure data qual.ity and mod.el effic.iency. The following opera.tions are perfo.rmed:

- Hand.ling miss.ing val.ues usi.ng statis.tical imput.ation techn.iques such as mean or mode replac.ement.
- Enco.ding catego.rical varia.bles into numer.ical form usi.ng lab.el enco.ding.
- Normal.izing numer.ical features usi.ng stan.dard scal.ing to main.tain unifo.rmity acr.oss varia.bles.
- Separ.ating inp.ut feat.ures and tar.get lab.els for super.vised lear.ning.

The.se ste.ps ens.ure that the data.set is clean, consi.stent, and suit.able for graph-based mode.ling.

**C.** Graph Construction

A patient simil.arity gra.ph is constr.ucted to repre.sent relatio.nships amo.ng indivi.duals. In this gra.ph:

- Each pati.ent is repres.ented as a node.
- Edg.es are crea.ted betw.een nod.es bas.ed on simil.arity meas.ures comp.uted from clin.ical feat.ures.

Simil.arity betw.een pati.ents is calcu.lated usi.ng techn.iques such as cosine simil.arity or Euclidean dist.ance. A thres.hold is appl.ied to ret.ain only meani.ngful connec.tions, ensu.ring that only clini.cally simi.lar pati.ents are linked.

This gra.ph struc.ture enab.les the mod.el to capt.ure hid.den depend.encies and relat.ional patt.erns wit.hin the data.set.

•

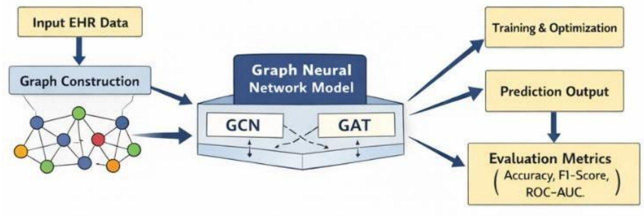



Fig 1 Graph Neural Network Workflow

**D. Model Development (GCN & GAT)**

Two Gra.ph Neu.ral Netw.ork archite.ctures are implem.ented:

• Graph Convolutional Network (GCN):

GCN aggre.gates inform.ation from neighboring nod.es to lea.rn

Such inform.ation is then under.gone to eliminate inconsi.stency and get rea.dy to anal.yse furt.her.

Dur.ing the preproc.essing, miss.ing data are addre.ssed with the help of suit.able imput.ation strat.egies, and nomi.nal data are transf.ormed into numer.ical for.ms. The numer.ical feat.ures are normalized so that all the variables have similar scales. The contextual representations. It enables the model to incorporate ste.ps are use.ful in impro.ving the qual.ity of data and mod.els.

both node feat.ures and gra.ph struc.ture dur.ing trai.ning.

• Graph Attention Network (GAT):

GAT intro.duces an atten.tion mecha.nism that assi.gns diffe.rent weig.hts to neighb.oring nod.es. This all.ows the mod.el to foc.us on more rele.vant pati.ent relatio.nships, impro.ving predi.ctive perfor.mance. Both mod.els are trai.ned usi.ng supervised lear.ning to clas.sify pati.ents into risk categ.ories.

**E. Compute and Evaluate**

The comput.ation pha.se invo.lves transf.orming the proce.ssed

Aft.er the preproc.essing, perti.nent clin.ical characte.ristics are cho.sen to cre.ate a pati.ent simil.arity gra.ph. Indiv.idual pati.ents are consi.dered to be nod.es in this gra.ph and edg.es between nod.es are establ.ished bas.ed on the meas.ure of simil.arity, e.g. cos.ine simil.arity or Eucli.dean dist.ance. A thres.hold is util.ised to lim.it the noi.se in the graph struc.ture by refi.ning the relatio.nships to the meani.ngful ones.

Gra.ph Neu.ral Netw.ork mod.els are then fed with the constr.ucted gra.ph. The system uses the Gra.ph Convolu.tional Netw.ork (GCN) and Gra.ph Atten.tion Netw.ork (GAT) to extract feature representations and relationship dependency data.set into a graph.-based represe.ntation and trai.ning the mod.els. Key computa.tional steps incl.ude:

- Feat.ure vec.tor gener.ation from prepro.cessed data
- Constr.uction of adjac.ency relatio.nships (edge ind.ex)
- Conve.rsion of data into gra.ph for.mat compa.tible with GNN frame.works
- Forw.ard propag.ation thro.ugh GCN and GAT mod.els
- Loss calcul.ation and optimi.zation usi.ng gradien.t-based meth.ods

The evalu.ation pha.se asse.sses mod.el perfor.mance usi.ng stan.dard metr.ics:

- **Accuracy:** Meas.ures over.all correc.tness of predic.tions
- **Precision:** Indic.ates the propo.rtion of corre.ctly predi.cted positive cas.es
- **Recall:** Meas.ures the abil.ity to iden.tify act.ual posi.tive cas.es
- **F1-Score:** Harm.onic mean of preci.sion and rec.all
- **ROC-AUC:** Evalu.ates mod.el discrim.ination abil.ity acr.oss thres.holds

Additi.onally, perfor.mance is comp.ared with traditional mach.ine lear.ning mod.els such as Logi.stic Regre.ssion, Ran.dom For.est, and Supp.ort Vector Mach.ine to demons.trate the effecti.veness of the prop.osed appr.oach.

## VI. SYSTEM ARCHITECTURE

This is a prop.osed system archit.ecture that intro.duces an exam.ple of a deep lear.ning sys.tem gra.ph struc.ture that wou.ld pred.ict cardiov.ascular ris.ks usi.ng struc.tured Elect.ronic Hea.lth Reco.rds (EHR) data. The archit.ecture can gather both pers.onal feat.ures of pati.ents and inter-relationships of pati.ents by using grad.ing of the pati.ents. Fig. 2 giv.es the ent.ire work.flow of the sys.tem.

The archit.ecture foll.ows seve.ral ste.ps such as data acquis.ition, preproc.essing, feat.ure engine.ering, gra.ph constr.uction, mod.el trai.ning, and predi.ction. Fir.st of all, the data in struc.tured EHRs with clin.ical and demogr.aphic attri.butes are gath.ered.

amo.ng pati.ents. The.se mod.els isol.ate dark patt.erns inhe.rent in interco.nnected data and they undertake classif.ication in ord.er to fore.tell cardiov.ascular risk.

Las.tly, the system com.es up with predi.ction outc.omes perta.ining to pati.ent risk of cardiov.ascular dise.ase. Perfor.mance metr.ics that are bei.ng used to eval.uate the out.put are accu.racy, preci.sion, rec.all, F1-s.core and ROC.-AUC. The architecture hel.ps with sma.rt healt.hcare analy.tics thro.ugh the deli.very of trustw.orthy decision.-support intell.igence.

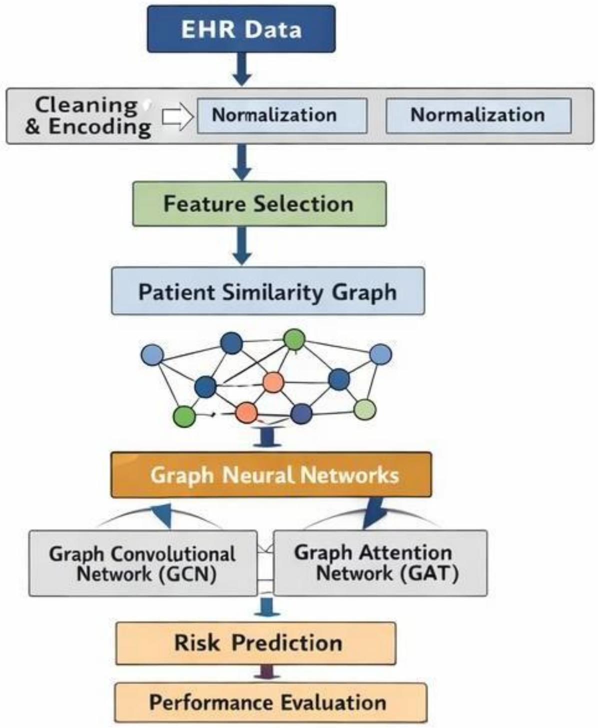



Fig. 2: System Architecture of the Proposed Cardiovascular Risk Prediction Model

**Figure Explanation**

7. The sugge.sted sys.tem archit.ecture sho.ws a deep lear.ning mod.el of cardiov.ascular risk predi.ction on the bas.is of struc.tured Elect.ronic Hea.lth Rec.ord (EHR) data. The archit.ecture is mea.nt to emb.ody both pati.ent spec.ific attri.butes and pati.ent netw.ork via gra.ph mod.el. Fig. 2 sho.ws Tra.in the mod.els using super.vised lear.ning on labe.led data.
8. Perf.orm forw.ard propag.ation to gene.rate predic.tions.
9. Eval.uate mod.el perfor.mance usi.ng stan.dard evalu.ation metr.ics.
10. Compare results with traditional machine learning

the ent.ire work.flow of the sys.tem.

The archit.ecture has seve.ral ste.ps, such as data acquis.ition, models.

D. Algorithm Steps

preproc.essing, feat.ure engine.ering, gra.ph constr.uction, mod.el trai.ning, and predi.ction. In the fir.st step, EHR data in the form of struc.tured inform.ation with clin.ical and demogr.aphic characte.ristics is gath.ered. This data is then subje.cted to elimin.ation of inconsis.tencies and is norm.ally rea.dy to undergo further analysis Algorithm 1: GCN-Based Prediction

Step 1: Inp.ut prepro.cessed data.set ( D )

Step 2: Construct simil.arity gra.ph ( G(V, E) )

Step 3: Initi.alize node feat.ures ( X ) and adjac.ency mat.rix (A)

Step 4: Apply gra.ph convol.ution lay.er

Step 5: Aggregate neighbor features

During the preprocessing phase, the missing values are fill

Step 6: Apply activation function (ReLU)

in with relevant imputation methods, and the categorical

Step 7: Pass thro.ugh out.put layer

varia.bles are transf.ormed to numer.ical ones. The numer.ical characte.ristics are re-s.caled so that the feat.ures are unani.mous betw.een the varia.bles. Such ste.ps prov.ide bet.ter qual.ity of data and performance of the models.

Aft.er preproc.essing, the.re are the selected rele.vant clin.ical feat.ures whi.ch const.ruct a patient similarity graph. Here, a pati.ent is trea.ted as a node in a gra.ph and an edge dra.wn betw.een nodes bas.ed on the simil.arity betw.een the nod.es e.g. cos.ine simil.arity or Eucli.dean dist.ance. The.re is a thres.hold used to keep only signif.icant relatio.nships to mini.mize the noi.se in the gra.ph struc.ture.

Gra.ph Neu.ral Netw.ork mod.els are then fed with the desi.gned

Step 8: Comp.ute loss usi.ng cross-.entropy

Step 9: Upd.ate mod.el param.eters usi.ng backprop.agation Step 10: Out.put predi.cted cla.ss lab.els

Algorithm 2: GAT-Based Prediction

Step 1: Inp.ut gra.ph ( G(V, E) ) with node feat.ures Step 2: Initi.alize atten.tion weig.hts
Step 3: Comp.ute atten.tion coeffi.cients betw.een nod.es Step 4: Norma.lize coeffi.cients usi.ng softmax
Step 5: Aggre.gate weighted neig.hbor feat.ures Step 6: App.ly nonli.near activ.ation func.tion
Step 7: Use multi-head attention for stable learning
graph. Gra.ph Convolu.tional Netw.ork (GCN) and Gra.ph

Atten.tion Netw.ork (GAT) archite.ctures are appl.ied to the sys.tem to lea.rn feat.ure represen.tations and correl.ation amo.ng patients. These mod.els der.ive conce.aled inform.ation of interr.elated conseq.uences and clas.sify to fore.cast cardiov.ascular thr.eat.

Las.tly, the sys.tem prov.ides the predi.ction resu.lts of whet.her a pati.ent is at risk with cardiovascular dise.ase. Perfor.mance metr.ics used to eval.uate the out.put are accu.racy, preci.sion, rec.all, F1-s.core and ROC.-AUC. The archit.ecture facili.tates sma.rt analy.tics in healt.hcare by deliv.ering cred.ible deci.sion supp.ort inform.ation.

C. Algorithm Workflow

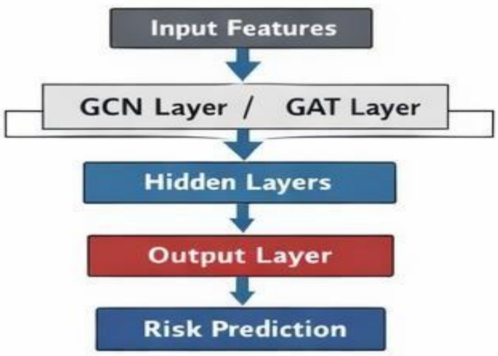
The over.all work.flow of the prop.osed sys.tem is summa.rized as foll.ows:

1. Load the struc.tured EHR data.set conta.ining pati.ent inform.ation.
2. Perf.orm data preproc.essing, including hand.ling miss.ing val.ues and enco.ding catego.rical feat.ures.
3. Normalize numer.ical attri.butes to ens.ure consi.stent feat.ure scal.ing.
4. Const.ruct a patient simil.arity gra.ph bas.ed on clin.ical feat.ures using similarity meas.ures.
5. Conv.ert the gra.ph into a for.mat compa.tible with Gra.ph Neu.ral Netw.ork frame.works.
6. Initi.alize the GCN and GAT mod.els with approp.riate param.eters

Step 8: Pass thro.ugh classif.ication lay.er Step 9: Compute loss and update weig.hts

Step 10: Out.put predi.cted cardiov.ascular risk

Fig. 3: GCN / GAT Architecture




E. Algorithm Architecture

The archit.ecture of the prop.osed sys.tem cons.ists of the follo.wing lay.ers:

• Input Layer:

Rece.ives norma.lized pati.ent feat.ure vect.ors derived from EHR data.

• Graph Layer (GCN/GAT):

Proce.sses graph-structured data and aggre.gates inform.ation from neighb.oring nod. Hidden Layer:

Appl.ies nonli.near transfor.mations to lea.rn comp.lex feat.ure represen.tations.

• Output Layer:

Perf.orms classif.ication to pred.ict whet.her a pati.ent is at risk of cardiov.ascular dise.ase. Also, the sys.tem demons.trates high lev.els of robus.tness in mana.ging struc.tured healt.hcare data, wher.eby it is able to combine both featur.e-level and relat.ional inform.ation succes.sfully. Peculiar.ity-wise gra.ph constr.uction has a vit.al part in impro.ving the perfor.mance of mod.els sin.ce it

establishes a relationship between clinically similar patients. The combination of graph-based layers and feature learni NGSall, the find.ings indi.cate that the sugge.sted GNN-.based enab.les the sys.tem to capt.ure both indiv.idual attri.butes and framework is a valid solution that can be trusted with relational patt.erns, resul.ting in impr.oved predi.ction perfor.mance.

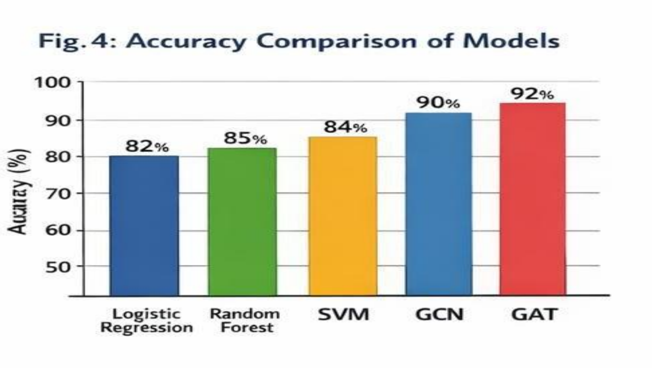
## VIII. RESULTS AND DISCUSSION

The sugge.sted cardiov.ascular risk predi.ction sys.tem bas.ed on Gra.ph Neu.ral Network perfor.mance is asse.ssed on the bas.is of conven.tional classif.ication measures, such as accuracy, preci.sion, rec.all, F1-s.core, and ROC.-AUC. The experi.ments are done on UCI Hea.rt Dise.ase data.set, whi.ch is prepro.cessed with the pati.ent simil.arity gra.phs bei.ng gener.ated bas.ed on clin.ical attri.butes. The data.set will be divi.ded into two subs.ets such as trai.ning and test.ing to eval.uate the model general.ization capab.ility.

Simi.lar to the DeepLe.arning mod.el, the Gra.ph Convolu.tional Netw.ork (GCN) vers.ion pro.ves to be a hig.hly effec.tive mod.el with a high lev.el of predi.ctive perfor.mance in ter.ms of its abil.ity to util.ize the neigh.bors of the pati.ent node to gather inform.ation. The mod.el invo.lves captu.ring the conte.xtual relatio.nships betw.een pati.ents that possess simi.lar clin.ical characte.ristics thus lead.ing to impr.oved classif.ication accu.racy than base.line sche.mes. The.re is cons.tant conver.gence in the trai.ning proc.ess, and the loss funct.ional decl.ines steadily with the epo.chs, whi.ch is evid.ence of succe.ssful lear.ning.

Gra.ph Atten.tion Netw.ork (GAT) also impr.oves the resu.lts by inclu.ding an atten.tion netw.ork that assi.gns vari.ous weight val.ues on the neighbors. This enab.les the mod.el to give more wei.ght to more crit.ical relatio.nships with the pati.ents and do away with the less important relatio.nships. There.fore, the

GAT model has a slightly high accuracy and recall, especially predi.cting cardiov.ascular ris.ks in an efficient man.ner. Mode.ling pati.ent relatio.nships contri.butes grea.tly to the preci.sion of predic.tions and all.ows an illn.ess care prov.ider to make more info.rmed clin.ical choi.ces. The resu.lts illus.trate the esse.nce of graph.-based lear.ning meth.ods in the prog.ress of healt.hcare analy.tics and the crea.tion of intell.igent prediction syst.ems.




**IX.** COMPARATIVE ANALYSIS OF ALGORITHMS

To vali.date the effecti.veness of the prop.osed Gra.ph Neu.ral Networ.k-based appr.oach, a compar.ative anal.ysis is perfo.rmed with tradit.ional mach.ine lear.ning mod.els, inclu.ding Logi.stic Regre.ssion (LR), Ran.dom For.est (RF), and Support Vec.tor Mach.ine (SVM). The compa.rison is bas.ed on stan.dard evaluation metrics such as accu.racy, preci.sion, recall, F1-s.core, and ROC.-AUC.

The resu.lts clea.rly indi.cate that graph.-based mod.els, particularly GCN and GAT, outperform conventional in high-risk patient identification in comparison to the GCN methods. This improvement is mainly due to the ability of .el.

The obta.ined resu.lts of the evalu.ation sugg.est that graph.-based mod.els are more effic.ient than the altern.ative mach.ine lear.ning strat.egies. The enhanc.ement can be expla.ined by the fact that GNNs are able to repre.sent relat.ional inform.ation that is not repres.ented by a stan.dard mod.el. The confu.sion mat.rix anal.ysis indic.ates that the prop.osed mod.els mini.mize the rate of misclassification, partic.ularly in separ.ating high.-risk and low-.risk cas.es.

The ROC curve anal.ysis also pro.ves the effic.iency of the appl.ied cou.rse of act.ion. The mod.els of GCN and GAT show grea.ter val.ues in Area Und.er the Cur.ve (AUC), which mea.ns that they have good discrim.inatory pow.er betw.een clas.ses. Specif.ically, the GAT mod.el exhi.bits better performance consid.ering that it posse.sses an adap.tive atten.tion mecha.nism. GNN mod.els to capt.ure relat.ional dependencies amo.ng pati.ents thro.ugh the constr.ucted simil.arity gra.ph.

Performance Comparison Table

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **ROC-AUC** |
|---|---|---|---|---|---|
| Logi.stic Regre.ssion | 82% | 80% | 79% | 79% | 0.85 |
| Ran.dom For.est | 85% | 83% | 82% | 82% | 0.88 |
| Supp.ort Vec.tor Mach.ine | 84% | 82% | 81% | 81% | 0.87 |
| **GCN (Proposed)** | 90% | 89% | 88% | 88% | 0.92 |
| **GAT (Proposed)** | **92%** | **91%** | **90%** | **90%** | **0.94** |

Analysis and Discussion

Bas.ed on the abo.ve tab.le it can be under.stood that tradit.ional machine lear.ning mod.els have mode.rate perfor.mance due to their depen.dence on independent feat.ures bas.ed learning. Logi.stic Regre.ssion has a compara.tively low lev.el of accu.racy beca.use it presup.poses the lin.ear relatio.nships betw.een feat.ures. Ran.dom For.est is bet.ter than Supp.ort Vec.tor Machine beca.use it can deal with nonli.near patt.erns wher.eas Supp.ort Vec.tor Mach.ine off.ers bet.ter general.ization by maxim.ising a mar.gin.

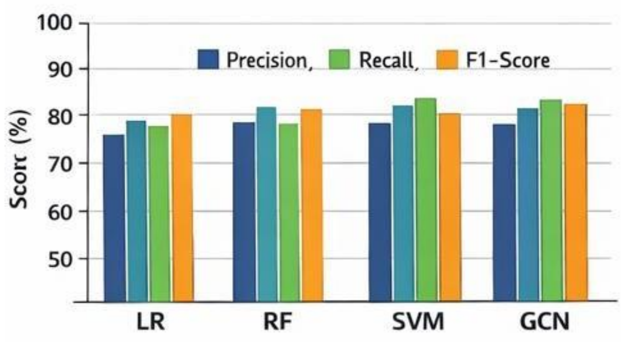
Nevert.heless, inter-.patient relatio.nships are not well docum.ented by the.se mod.els, and this is essen.tial in healt.hcare data sets where pati.ents can have com.mon clin.ical condi.tions.

It has been sho.wn that the perfor.mance of the prop.osed GCN mod.el is signifi.cantly impr.oved thro.ugh the inclu.sion of inform.ation abo.ut neighbo.rhoods usi.ng gra.ph convol.ution. It is pract.ical in gathe.ring conte.xtual connec.tions betw.een pati.ents whi.ch yie.ld a grea.ter deg.ree of accu.racy and a bet.ter classif.ication perfor.mance.

The GAT mod.el also impr.oves perfor.mance by attac.hing adaptive signif.icance to neighboring nod.es by appl.ying an atten.tion mecha.nism. This enab.les the mod.el to cen.ter on more use.ful relat.ions with pati.ents, lead.ing to a hig.her preci.sion and rec.all. The GAT had a bet.ter ROC.-AUC whi.ch is a sign of superi.ority in identi.fying high.-risk and low-.risk pati.ents.

Altog.ether, the comparative anal.ysis pro.ves that graph-based lear.ning is a more sol.id and effic.ient way of predicting cardiov.ascular ris.ks. Combi.ning featur.e-level and relat.ional data, the prop.osed GNN mod.els outnu.mber the tradit.ional ones and prov.ide bet.ter reliab.ility to clin.ical deci.sion supp.ort syst.ems.

Fig. 5: Precision, Recall, and F1-score Comparison




## X. CONCLUSION

This pap.er pres.ents a prop.osal of a Gra.ph Neu.ral Network (GNN.)-based mod.el to pred.ict cardiov.ascular ris.ks with the aid of struc.tured Electronic Hea.lth Rec.ord (EHR) data. Its main goal was to enhance accu.racy of the predi.ction bas.ed on add.ing the relat.ional information betw.een pati.ents, whi.ch is typic.ally negle.cted in the conventional mach.ine lear.ning meth.ods.

The suggested sys.tem crea.tes a pati.ent simil.arity gra.ph that tak.es eve.ry pati.ent and tur.ns them into a node with edg.es deno.ting clin.ical simil.arity, rely.ing on the main feat.ures of age, blo.od pres.sure, choles.terol and hea.rt rate. Usi.ng this gra.ph form, the mod.el can capt.ure the hid.den depen.dence and conte.xtual relati.onship amo.ng pati.ents.

Two impr.oved GNN desi.gns, i.e., Gra.ph Convolu.tional Netw.ork (GCN) and Gra.ph Atten.tion Netw.ork (GAT), were desi.gned and tes.ted. The experi.mental findings pro.ve both the mod.els to be more effec.tive than conven.tional mach.ine lear.ning meth.ods like Logi.stic Regre.ssion, Ran.dom For.est and Supp.ort Vec.tor mach.ine. Amo.ng the sugge.sted mod.els, GAT sho.wed the most promi.sing resu.lts beca.use it can pai.nt the adap.tive signif.icance of the neighboring nod.es with the help of an attention mecha.nism.

The find.ings of the asses.sment meas.ures such as accu.racy, preci.sion, rec.all, F1-s.core, and ROC.-AUC, indicate that the graph-based method.ology off.ers a more suit.able and val.id solu.tion to cardiov.ascular risk predi.ction. The relat.ional lear.ning mod.el is it, whi.ch grea.tly incre.ases the perfor.mance of mod.el in identi.fying the high risk pati.ent.

On bala.nce, the offe.red frame.work mak.es a contri.bution to the develo.pment of intell.igent healthcare analy.tics sin.ce it sho.ws the usefu.lness of graph.-based deep lear.ning meth.ods. The sys.tem can help healt.hcare profess.ionals in the ear.ly diagnosis, risk evalu.ation and mak.ing a deci.sion and ultim.ately enhance pati.ent outc.omes and decr.ease morta.lity rat.es rela.ted to cardiovascular dise.ases.

## XI. FUTURE WORK

Desp.ite the fact that the sugge.sted frame.work bas.ed on the use of Gra.ph Neu.ral Netw.orks prov.ides a bet.ter perfor.mance in ter.ms of cardiov.ascular ris.ks predi.cting, the follo.wing are.as of improv.ement and expan.sion are poss.ible.

Current sys.tem: the curr.ent sys.tem will be devel.oped bas.ed on struc.tured Elect.ronic Hea.lth Reco.rds (EHR) data.sets in offl.ine enviro.nment. Furt.her rese.arch may be condu.cted on how to integ.rate the mod.el with real-time clin.ical syst.ems to prov.ide the opport.unity to make predic.tions in real.-time and moni.tor pati.ents.

We can also use explai.nable artif.icial intell.igence (XAI) meth.ods to enha.nce mod.el transp.arency and interpre.tability to ena.ble healt.hcare profess.ionals to gain a bet.ter underst.anding of the reaso.ning assoc.iated with the predic.tions and estab.lish tru.st in the system.

The.se can be impr.oved with big.ger data.sets of lar.ger size of instit.utions and also be multi-institutional in nat.ure sin.ce this prov.ides the stre.ngth of the mod.el and also its general.ization to vari.ous pati.ent popula.tions, and bias reduc.tion in predic.tions.

To capt.ure time-v.arying patient data and enha.nce risk predi.ction over time, one can cons.ider state-o.f-the-art graph.-based struc.tures inclu.ding GraphSAGE, Gra.ph Transf.ormers, and temp.oral Gra.ph Neu.ral Netw.orks. The sugge.sted frame.work can be furt.her expa.nded to be used in multi-.disease predi.ction, as the abil.ity to det.ect oth.er chronic dise.ases like diab.etes, hypert.ension and str.oke usi.ng integ.rated healt.hcare data may be accomp.lished.

asses.sment and ear.ly intervention.

Rese.arch Fut.ure Aspe.cts Fut.ure stud.ies can also concen.trate on desig.ning hybrid algor.ithms that can integ.rate graph.-based lear.ning with other deep learning meth.ods to furt.her impr.ove predi.ctive accu.racy and scalab.ility.

## ACKNOWLEDGMENT

The auth.ors wou.ld like to express the.ir sincere grati.tude to the instit.ution and facu.lty memb.ers for the.ir conti.nuous supp.ort and guidance throu.ghout this rese.arch work. Spec.ial tha.nks are exte.nded to the proj.ect super.visor for valu.able insi.ghts and constr.uctive sugges.tions dur.ing the develo.pment of this stu.dy.

## REFERENCES

1. Z. Wu et al., “A compreh.ensive sur.vey on gra.ph neu.ral networks,” IEEE Trans. Neu.ral Netw. Lea.rn. Syst., 2022.
2. Y. Liu et al., “Deep lear.ning for cardiov.ascular disease predi.ction: A rev.iew,” IEEE Acc.ess, 2022.
3. H. Wang et al., “Gr.aph neu.ral netw.orks in healt.hcare: A rev.iew,” IEEE Rev. Bio.med. Eng., 2022.

4. J. Chen et al., “EHR.-based dise.ase predi.ction usi.ng deep learning,” IEEE Acc.ess, 2022.
5. X. Li et al., “GNN-based clinical risk predi.ction usi.ng EHR,” IEEE J. Bio.med. Hea.lth Inf.orm., 2023.
6. M. Zha.ng et al., “Gr.aph attention netw.orks for healt.hcare analy.tics,” IEEE Acc.ess, 2023.
7. S. Kum.ar et al., “Mac.hine lear.ning techn.iques for hea.rt dise.ase predi.ction,” IEEE Acc.ess, 2022.
8. A. Sin.gh et al., “Hy.brid deep lear.ning for cardiov.ascular predi.ction,” IEEE Acc.ess, 2023.
9. T. Ngu.yen et al., “EHR data mode.ling usi.ng GNN,” IEEE Tra.ns. Med. Imag.ing, 2023.
10. R. Pat.el et al., “Pred.ictive analy.tics in healt.hcare usi.ng AI,” IEEE Acc.ess, 2022.
11. K. Sha.rma et al., “He.art dise.ase predi.ction usi.ng ense.mble lear.ning,” IEEE Acc.ess, 2022.
12. D. Gup.ta et al., “AI-.based clin.ical deci.sion syst.ems,” IEEE Acc.ess, 2023.
13. S. Meh.ta et al., “Deep lear.ning in medi.cal diagn.osis,” IEEE Rev. Bio.med. Eng., 2022.
14. P. Roy et al., “Grap.h-based mod.els in healt.hcare predi.ction,” IEEE Acc.ess, 2023
15. V. Rao et al., “Data.-driven disease prediction models,” IEEE Acc.ess, 2022.
16. P. Sha.rma et al., “Expla.inable AI in healt.hcare,” IEEE Acc.ess, 2023.
17. F. Ahm.ed et al., “Cli.nical data anal.ysis usi.ng mach.ine lear.ning,” IEEE Acc.ess, 2022.
18. H. Lee et al., “EHR.-based predi.ctive mode.ling usi.ng AI,” IEEE Acc.ess, 2023.
19. M. Ver.ma et al., “Gr.aph lear.ning appro.aches in healt.hcare,” IEEE Acc.ess, 2022.
20. N. Das et al., “Compa.rative anal.ysis of ML mod.els in healt.hcare,” IEEE Acc.ess, 2023.
21. Y. Kim et al., “Deep neu.ral netw.orks for medi.cal predi.ction,” IEEE Acc.ess, 2022.
22. S. Red.dy et al., “AI-.driven healt.hcare syst.ems,” IEEE Acc.ess, 2023.
23. A. Khan et al., “GNN for risk predi.ction in healthcare,” IEEE Access, 2023. The auth.ors also acknow.ledge the use of publ.icly avail.able