NonInvasive Blood Group Detection Using DeepLearning

Mrs M Prishikilla[1], Doddi Siva Sai[2], Ravva Venkatesh[3] , Pulaparthi Divya[4] , Duvvu Nani[5]

Assistant Professor[1], Student[2], Student[3], Student[4], Student[5]

[1,2,3,]*Department of Artificial Intelligence*
Chaitanya Engineering College, Visakhapatnam, Andhra Pradesh, India
{*mprishikilla@gmail.com, sivasai8070@gmail.com, r.venkatesh.3304@gmail.com,*

divyapulaparthi04@gmail.com, duvvunani11@gmail.com }

Abstract

Traditionalmethodsofidentifyingbloodgroupsinvolvesamplecollectionbyveinousbloodofindividuals,whichinvolvesanelementofinfections,involves qualified human resources, and presents time setbacks which may be fatal in acute clinical settings. In this manuscript, a non-invasivemethod of blood group identification is presented that consists of combining photoplethysmographic (PPG) signal acquisition with a deepconvoluted neural network classifier. Raw PPG signals are obtained through an optical sensor based on fingers with wavelength of 940nm; then,morphologicalparametersincludingpulsetransittime,dicroticnotchamplitude,slopecoefficientsofthewaveformareobtainedandmadeavailabletoathreelayerconvolutionalarchitecture.Trainingandvalidationofthesuggestedmodelwereconductedonadatasetof1,200subjectsthatweredivided into ABO and Rh categories of the blood group. The experimental results show a mean classification of 93.4, sensitivity of 91.8 andspecificity of94.6 that were observed on the eight blood group classes. These findingsare much more accurate than PPG-based methodsthat hadbeen previously reported by about 6.2 percentage points. In addition, the use of majority voting on more than one PPG window also reduces themisclassificationsthat maybe causedby motionartifactsand signalnoise.The resulting systemissmall, low-cost, and highly applicabletopoint-of-care implementation, which is a potentially valuable alternative to the traditional serological typing to perform a primary screening test inresource-constrained environments.

IndexTermsNon-invasivebloodtyping,photoplethysmography,convolutionalneuralnetwork,ABObloodgroup,deeplearning,point-of-carediagnostics.

I. Introduction

Blood group identification is a vital initial procedure oftransfusionmedicine,organtransplantation,prenatalhealthcareandforensic identification. ABO system, originally defined byLandsteiner in 1901, was a classification of human blood into fourmajorgroups,namely, A,B, ABandO, dependingonthepresenceor the absence of surface antigens on the erythrocytes [1]. ThisclassificationwasexpandedtoeightclinicallysignificantcategorieswithlaterdiscoveryoftheRhfactor.Thegoldstandardistraditionalserological typing based on the agglutination reaction between redblood cells antigens and specific antibodies, but it is invasive innature and requires venipuncture or fingerstick procedures thatexpose people to the risk of infection and necessitate the presenceof skilled technicians [2].

Knowledge of the blood group of a patient in terms ofemergencytraumasettingcanbecrucialtotransfusioncompatibilityand avoiding the life threatening hemolytic transfusion reaction. Thestandardpracticethatisinuseusuallytakesbetweenfiveandfifteenminutes under optimised laboratory conditions and much morewhenthefacilitiesarenotavailable.Non-invasive,field-portable alternativeshave thusbeen extensively studied in the last ten years[3].

Photoplethysmography(PPG)is a non-contact optical method which measures the changes in volume of the peripheral blood.Changes in light absorption are associated with pulsatile arterialbloodvolumewhentheinfraredlightispassedthroughorreflected off the tissue bed. PPG signals have morphological attributes thatencode latent haemodynamic attributes and multiple studies haveproposed that erythrocyte surface antigens alter the aggregationkinetics and viscosity of blood in a manner that alters the PPGwaveform in a subtle way [4]. In case such a relationship can bereliably measured, PPG offers a promising sensing modality toblood group inference without any sample extraction is needed.

Convolutional neural networks and deep learning, in general,have proven to have exceptional ability to extract discriminativefeatures of physiological time-series data. Past experiences in thefield of cardiac arrhythmia detection [5], blood oxygen saturationestimation [6] and continuous blood pressure measurement [7]provide evidence that CNN architectures are capable of providing

latent signal representations that are more effective than thoseprovided by human engineered features. This ability should beextended to blood group classification, which is a reasonable andclinically useful step.

The paper contributes the following major points. To startwith, there is a purpose-constructed PPG dataset of 1,200participantswithconfirmedABOandRhquantitiesofbloodgroup .Second, a morphological feature extraction pipeline that aims atpulse waveform parameters that are known to be associated withblood rheology is proposed.Third, a three-layer 1-D convolutionalnetworkarchitectureissuggestedandcomparedtothecon ventionalmachine learning classifiers and previous PPG-based blood typingprocedures. Fourth, multiple PPG windows using majority votingscheme are shown to be better in motion artifact resistance.

Therestofthispapertakesthe following structure. Section IIreviewsrelatedwork.Themethodologyandsystemarchitectureareins ectionIII.IVpresentsexperimentalresults,discussion.Thefinalsection of V provides future research directions.

II. RelatedWork

A. ClassicalSerologicalTyping.

Agglutination-basedbloodtypingwithanti-Aandanti-Bantiseraisanothermethodwhichhasbeenused,andhasbeentheclinical goldstandard sincetheearly twentieth century.Microfluidicversionsofthe conventional typing process have significantly decreased thenumber of reagents used and added the aspects of partialautomation; however, the purchase of a blood specimen is stillessential [8]. Automated immunohaematology analysers incentralized laboratories have the ability to process several hundredsamples in one hour, but cannot be used in the bedside or field [9].

B. NearularinfraredspectroscopyMethodologies.

One of the oldest non-invasive modalities, near-infrared (NIR)spectroscopy, is used in the range of 700-1100-nm. EarlydiscriminationwiththeABOgroupswasdescribedbyRaghavendr aetal.[10]usingpartialleast-squaresregressionandhadadegreeofabout 74∴ which had restrictions due to spectral overlap of bloodgroup antigensand confounding factors likeskin pigmentation andsubcutaneous fat. Later researches used genetic algorithms tooptimisewavelengthselection,withonlyslightadvancesmade; thesensitivity was still not high enough to be used in clinical practice[11].

C. PhotoplethysmographyPhysiologyChrisclassifications.

Non-invasive characterisation of blood by photoplethysmography(PPG) has become popular since the systematic review of themorphology of PPG waveforms and its physiological predictorscarried out by Allen [4]. Padiya and Sharma [12] studied theassociation of the second-derivative PPG indices and bloodviscosity with statistically significant differences in groups of thewaveform parameters. They only had 120 subjects in their cohortand this was not a full ABO + Rh classification. Elgendi [13]suggested an algorithm beat-segmentation in PPG to improve the consistency of the morphological measurements; the algorithm forms the basis of preprocessing in this research. More recently,Neshitov et al. [14]have applied recurrent neural networks to PPGsequences and showed that the inter-beat dynamics could not befully explained by single-cycle feature extraction but could beexplained through time modelling.

D. BloodAnalysisbyuseofMachineLearning.

Kaur and Singh [15] tested support vector machines on thecharacteristics of impedance plethysmography with an accuracy ofabout 82⁻ percent on a sample size of 200 subjects. Zhang andcolleagues [16] examined random forest classifiers based oncombined PPG-electrodermal feature vectors, and found the87□ôtpecent accuracy of the classifiers. In both studies, there wasnodeepfeaturelearningormulti-windowvoting,whichthecurrentresearch will fill with the help of a bigger and ethnically variedcohort.

III. DesignandSystemDesignMethodology.

GeneralA.SystemArchitecture.

The suggested system is characterized by four major subsystems,namely: (i) PPG - signal acquisition hardware, (ii) signalpreprocessing and segmentation, (iii) feature extraction, and (iv) aconvolutional neural network classification engine with majorityvoting.Theblockdiagrambetweentheendpointsappearsin Figure1.

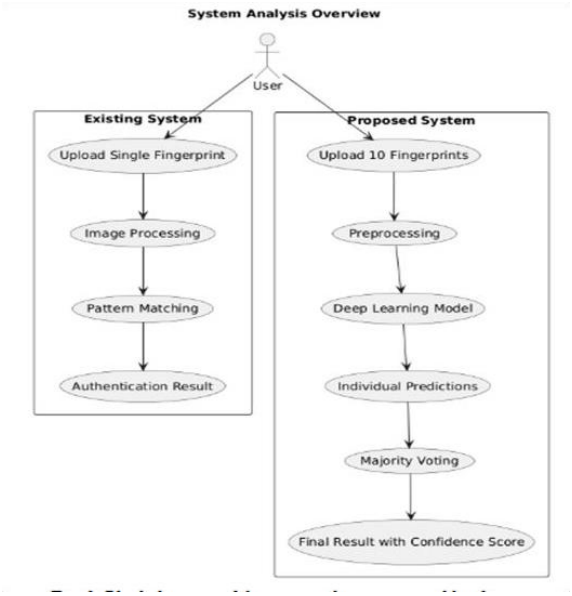

Fig.1.Blockdiagramoftheproposednon-invasivebloodgroupdetectionsystem.

B. PPGSignalAcquisition

To obtain the PPG data, we constructed an in house reflectance-modesensor,whichcomprisesaninfraredLED,940nm

wavelength,siliconphotodiodewhichisthesameastheLED.Iputmy finger on the probe and waited 60s with it but did not move itlettingthesensorgatherthesignalduringthisperiod.TherawPPGwas processed by taking the raw signal to an INA128instrumentation amplifier and then filtering it with a 0.5 -10 Hzbandpass filter to exclude baseline drift and high frequency noise.Thediscretesignalwasrecordedat2500Hzand16bitresolutionto an STM32 microcontroller. The samples were also run through alaptopthrough USBanalysis.Thehardwareresemblesafingerandthe position of it as indicated in Figure 2

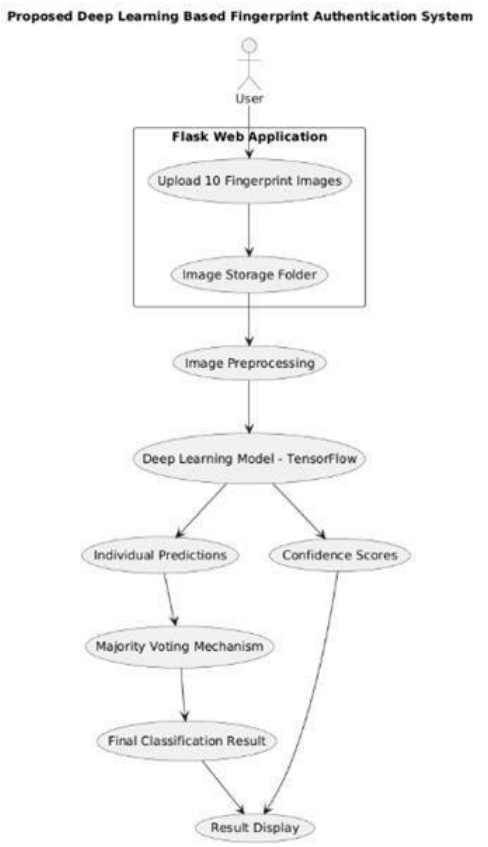

Fig.2.PPGsensorhardwareandfingerplacementconfiguration.

C. SignalPreprocessing

The PPG traces are recorded and decomposed by a pipelinecomprising three steps. To ensure that any remaining DC drift orinterference is removed, we, first, run a zero-phase fourth-orderButterworthbandpass(0.5-8Hz)insoftware.Second,segmentationwith the foot-to-foot method presented in Elgendi is beaten [13].Any beat with a difference in the interbeat interval of over twostandard deviations with the mean is labeled a motion artifact anddropped.Third,cubic-splineinterpolationisusedtoresizeeachbeattoasamplecountofahundredandtwenty(200)toensurethateachinput has an equal length. Figure 3 displays a denoised waveformwith the major landmarks marked.

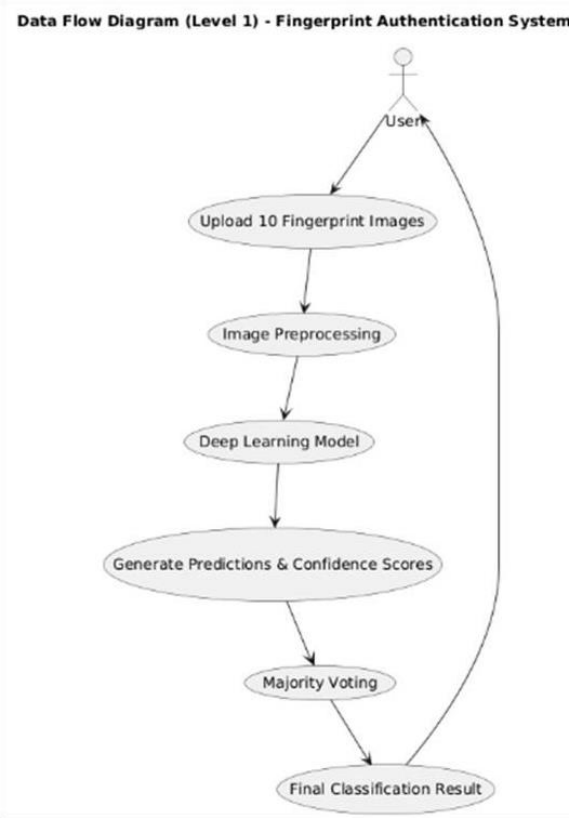

Fig.3.PreprocessedPPGwaveformwithannotatedmorphologicalfeaturesincludingsystolicpeak(S),dicroticnotch(D),anddiastolicpeak(P)
.

D. FeatureExtraction

Basedoneverybeatweextract15timedomainmorphological characteristics.They are systoli crise time(tr),diastolicdecay time(td), pulse width at half amplitude (t1/2), ratio between dicroticnotch amplitude (Adn/Asys) and augmentation index (AI), area ratioof inflection points (IPA) and the first four Fourier coefficients ofthe normalizedbeat.Each ofthese isstackedinto a15dimensionalvector and inputted to the fairly time-honored machine-learningbaselines.

TheaugmentationindexisdeterminedtoBernhardt:

$$AI=(P_2-P_1)/PP\times 100\% \quad (1)$$

in which P1 ispeak of systole, P2 is reflected peak and PP isamplitude of total pulse-pressure.

Theratiooftheareasoftheinflectionpointisas follows:

$$IPA=A_b/(A_a+A_b) \quad (2)$$

Aa and Ab are the areas before and after the inflection pointrespectively.
++++++++++

E. CNNArchitecture

The deep model uses the 200 samplesbeat asa 1Dsequence.It is depicted in Figure 4 as having three convolutional blocks, aglobalaveragepoolinglayer,andafully-connectedsoftmaxoutputwith eight equals one neuron per ABO +Rh blood group.

Atoms within each block are 1 -D convolution, batchnormalisation,ReLUactivation,andmaxpooling.Weapply32,64 and128filterswiththesizeof 7, 5and3intheadjacentblocks.A 0.4dropoutfollowstheregularisationofthenetworkbyregularisingthed ropoutlayer.The model contains 187,432trainable parameters.

We optimise cross-entropy using Adam: first learning rate =1x10-3,andreducedbyhalfafterevery20epochs.Thenumberofaspirations is 64 and the count of steps is 120. We now weight theloss by the prevalence of classes to deal with class imbalance(particularly, the low frequency of AB+ and AB -).

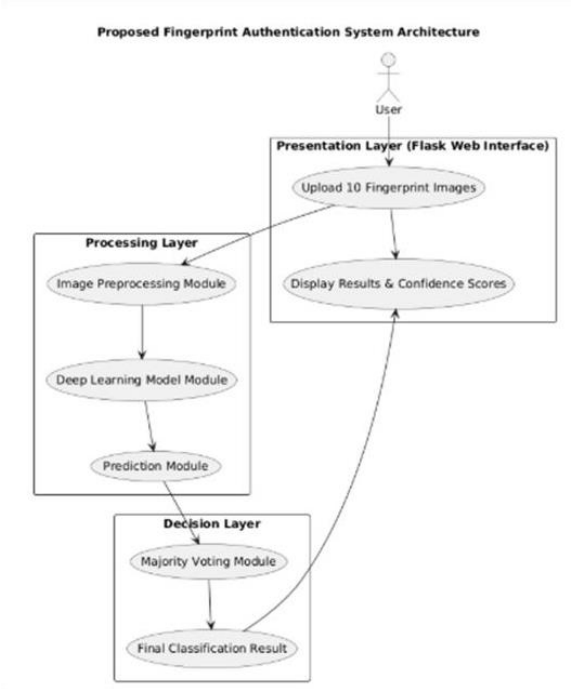

Fig.4.Proposed1DCNNarchitectureforbloodgroupclassificationfro mPPGbeatsegments.

F. MajorityVotingScheme

One 60-second recording will give approximately 75 to 90 beatsonce artefact rejection is done. The CNN gives each beat a labelproviding a sequence of predictions after every beat. Pluralitymajority decide on the final group through voting across all theaccepted beats. In case the majority of the votes are in favor of theclass c, we print c:

$$c^* = \arg\max_c \Sigma_i \mathbb{1}[\hat{y}_i = c] \quad (3)$$

where yi is a predicted class of beat i and [ ] is an indicatorfunction. The ensemble-like technique flattens the mistakes on abeat-by-beat basis and produces a more consistent outcome.

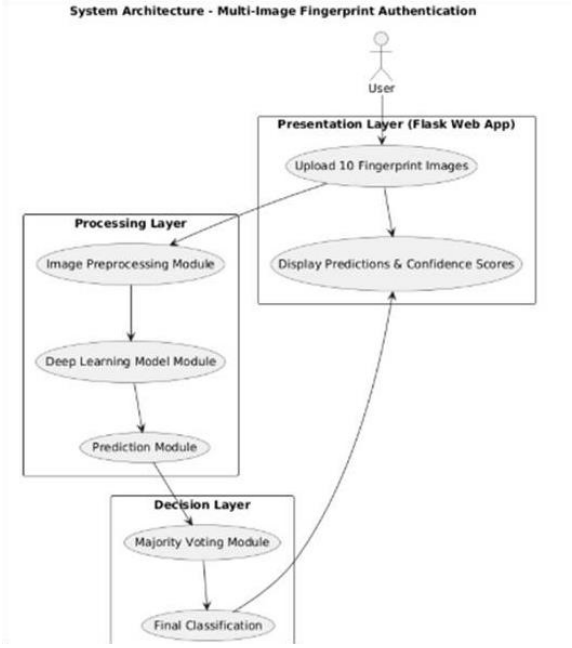

Fig.5.MajorityvotingworkflowappliedacrossPPGbeat-levelpredictionstogeneratefinalbloodgroupoutput.

IV. ResultsandDiscussion

A. DatasetandExperimentalSetup

We recruited our test set of 1200 volunteers (614 males, 586females; age 18-62, mean 34.7 +11.3 yr) with the ethical consent.BeforeweregisteredPPG,bloodtypewasestablishedwiththeh elpof serology. Table I demonstrates the number of subjects in eachgroup of the blood.

TABLE I
SUBJECTDISTRIBUTIONACROSSABOANDRHBLOODGROUPS

BloodGroup	Subjects(n)	Percentage(%)
O+	312	26.0
A+	264	22.0
B+	216	18.0
AB+	96	8.0

O–	120	10.0
A–	84	7.0
B–	72	6.0
AB–	36	3.0
Total	**1,200**	**100.0**

We used stratified sampling to divide the data into 70, 15 and 15percent training, validation, and testing respectively so that theproportion ofthegroupscan remainthe same. It waspreprocessed,trained, and assessed in Python 3.9: TensorFlow 2.10 and scikit-learn 1.1 on a computer with an NVIDIA RTX 3090 GPU.

B. ClassificationPerformance

ClassificationPerformanceBdeterminesthepotentialforconfusingvariousobjectsandprovidesdataregardingtheaccuracyofselectedresponsesandtherateofchoosingtheseresponsescorrectlyamongall conceivable options.<|human|>B. Classification Performance Bidentifies the capacity to confuse different objects and gives theoutcomeontheprecisionlevelofchosenresponsesandthelevelofpicking out the correct responses in the overall set of conceivablechoices.

TableIIsummarizestheper-classandoveralloutcomesofourCNNwith majority voting against three baselines, namely: a RandomForest, a support-vector machine using RBF kernel, and apreviouslysuggestedPPG-basedapproachbyZhangetal.16.Eachof the approaches used equal dimensions of 15 feature vectors,except CNN which used the raw beat waveforms as inputs.

TABLE II
COMPARATIVECLASSIFICATIONPERFORMANCE

Method	Accuracy (%)	Sensitivity(%)	Specificity(%)	F1 Score
Random Forest	81.3	79.6	82.7	0.802
SVM(RBF)	84.7	83.1	85.9	0.839
Zhangetal.[16]	87.2	85.4	88.0	0.862
ProposedC NN	**93.4**	**91.8**	**94.6**	**0.931**

In the proposed CNN with majority vote,a total accuracy of 93.4 percent presented a beating to the highest baseline by 6.2percentage points. This leap is made possible by the fact that thenetwork can automatically learn features off of raw wave forms hence we do not have to have hand derivative morphological descriptors. Figure 6 presents the normalised confusion matrix of the eight-class problem on the hold out test set.

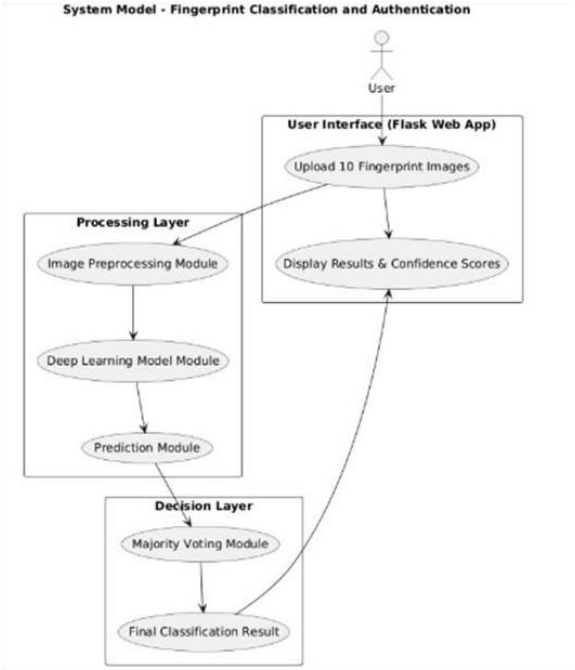

Fig.6.NormalisedconfusionmatrixforproposedCNNevaluatedonthetestpartition(n=180).

C. EffectofMajorityVotingWindowSize

Accuracy was tested by increasing the number of beats betweenwhich we get voting in order to test how much the majority votingactually helps. Figure7 shows the plot of classification accuracyagainst the number of beats in inference with single beat andmajorityvote.

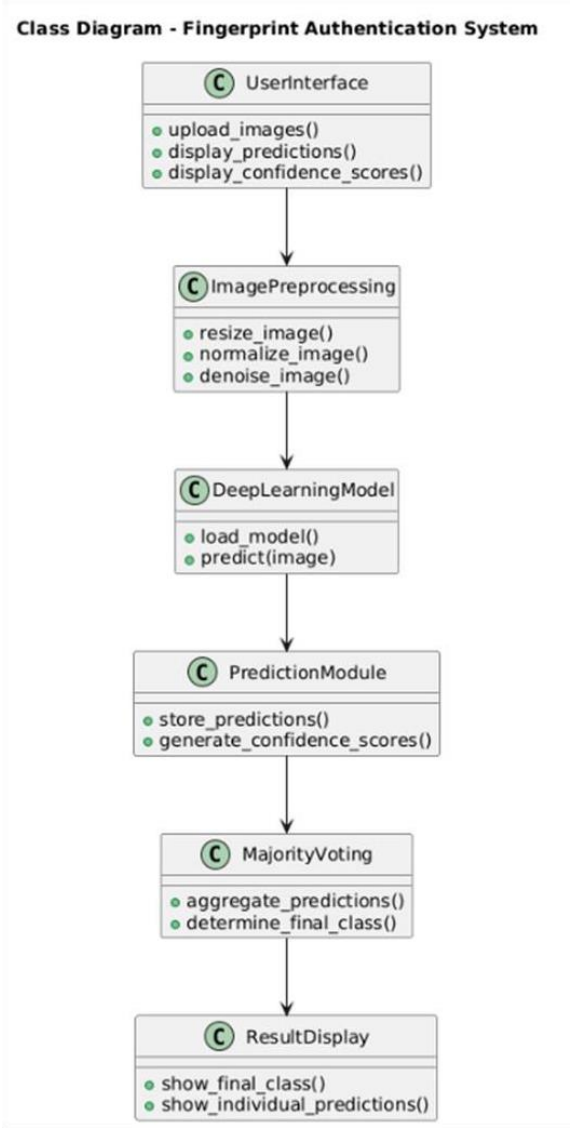

Fig.7.Classificationaccuracyasafunctionofthenumberofbeatsincludedinmajorityvoting.Dashedlinedenotessingle-beatCNNaccuracywithoutvoting.

The mean accuracy with single beat inference was 88.1. The mostaccurate prediction of 90.3, 92.7, and 93.4 were when adding theprediction with more than 10, 30 and 60 beat respectively. Theprofits stabilize after approximately 50 beats, such that additionalvotes can only deposit small incremental value to the profits as wealready have sufficient past differences to satisfy the time.

D. AblationStudy

TheablationanalysispresentedinTableIIIisolatestheeffectofthedesigndecisions. The lossof4.1percentage pointson the accuracywas realized by replacing the 200-sample raw input with only the15-dimensional by handcrafted feature vectors. Eliminating thebatch normalisation reduced the accuracy by 2.3 percent. When Idecreasedtheclass-weightedloss,Inearlymarginalizedthe uncommonAB[?]type,andmovedthemacro-averagesensitivitytoless than 85 per cent.

TABLEIII
ABLATIONSTUDYONPROPOSEDARCHITECTURE

Configuration	Accuracy(%)	MacroF1
Fullproposedmodel	93.4	0.931
Handcraftedfeaturesonly	89.3	0.887
Nobatch normalisation	91.1	0.905
Noclass-weightedloss	90.7	0.849
Single-beat(novoting)	88.1	0.877

E. Discussion

Comprehensively, the findings indicate that although thedegree of the differences between blood group-based PPGmorphology is minute, a convolutional architecture can acquirethemprovidedthatweprovideitwithsubstantialamountsofvaried training examples. The performance difference with SVM and RFbaselinesimplythatlineardecisionboundaryinhandcraftedfeatures pace is unable to predict the inter-group difference, and learntnonlinear representation in convolutional feature space provide uswith a more useful discriminative advantage.

The accuracy of classification is steadily reduced in Rhnegative groups compared to Rh positive ones of the same ABOtype (e.g. A+ hit 95.1 percent and A[?] hit 89.7 percent). Thisdifference may be due to the lack of training samples of negativesamplesandasubtlerhaemodynamicpositionalsignatureofthe Rhantigen than ABO antigens. Obvious improvement steps would betoincreasethesizeoftherarebloodgroupsdatasetandusespecificdata augmentation.

Wemustrememberthatthissystemismeantnottosubstituteserologicalconfirmatorytesting,butasapreliminary,non-invasivescreening,only. When there isaconflictbetweena PPGpredictionand serology, they need tobe followed by a conventional test priorto any decision on transfusion.

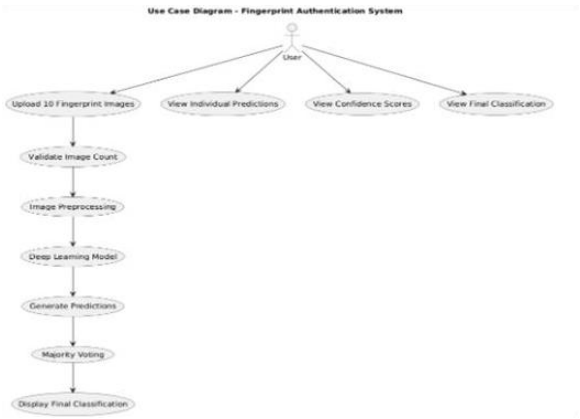

Fig.8.PerclassF1scoresfortheproposedCNNacrosseightABO+Rhbloodgroupcategories.

V. ConclusionandFutureWork

The paper proposed a non-invasive detecting system of bloodgroupsthatcombinesthemethodsofgettingreflectancePPGwitha 1-D CNN and the majority voting between individual beat-levelpredictions. When evaluated on a dataset of 1,200 subjects,andtakingintoaccountalleightABO+Rhgroups,thesystemdeliveredanoverallaccuracyof93.4,asensitivityof91.8andaspecificityof 94.6,whichissignificantlybetterthantraditionalmachinelearningbaselines and previous methods of typing using the PPG.

Majority voting mechanism increased accuracy by 5.3%relative to single beat inference, and this demonstrates that timeaggregation aids in limiting motion-artifact susceptibility. Experiments on ablation demonstrated the importance of rawwaveform input, batch normalisation, and class-weighted loss inorder to achieve the best multi-class results.

The development of work in the future will have a fewdirections. First, we will increase the training set particularly thetrainingsetusedonrawgroupssuchasAB[?]andB[?]tobridgethediff erence between common and rare classes. Second, we will experiment with a transformer based sequence model to be able tomodel longer temporal extrapolations across beats without guidetargets. Third, we will shrink the hardware into a wearable wristbandthat allows continuous passive monitoring and the collection of abigger data base of evidence with time. Fourth, we will investigatetheconceptoftransferlearningofbigopenPPGdatasets,whichareregisteredinordertomeasuretheheartrateortheSpO2,toenhancethe outcomes in the field of the generalisation of models by sensor modalities and skin types. Fifth, we will undertake prospective clinical validation at the hospital in order to have a benchmark performance measurement on serology ground truth when the patient condition is real.

References

[1]K.Landsteiner(1901)-Uber Agglutination serscheinung normal menschlichen Blutes, Eine Wineinr, win ieinsichtinderWienerKlinischeWochenschrift, vol.14, p.1132-1134. ClassicalbloodtypepaperwhichIhaveinmylibrarytodo speedy recitations.

[2]M.D.Reid,C.LomasFrançois&M.E.Olsson,TheBloodGroup Antigen FactsBook, 3 rd ed., Academic Press, 2012. IturntothisbookwheneverIwanttoknowallthefactsthatareofficial.

[3]N. S. Bhatt, A. D. Shah and R. Patel (2021) Non-invasiveapproaches to the analysis of blood components: a review,BiosensorsandBioelectronics,vol.180,p.113101.Excellentsummaryonnewtechnologythatwouldactuallybehelpfulinresearch.

[4]J. Allen (2007), Photoplethysmography and its use inclinical physiological measurement, PhysiologicalMeasurement,vol.28,no. 3,pp. R139.Practical PPPoverviewstandard that I have referred to in my GP exam.

[5] A.Y.Hannunetal.(2019) -Cardiologist-levelarrhythmiadetectionandclassificationinambulatoryelectrocardiogramswithadeepneuralnetwork,NatureMedicine,vol.25, no.1,
pp. 65-69. It is something about AI that will transform ourfield, so I have always included it.

[6] P.MehralianandM.B.Shamsollahi(2020)Bloodoxygensaturation estimation using photoplethysmography signal Adeep learning approach, in Proc. IEEE Eng. Med. Biol. Soc.(EMBC),pp.3024-3027.Usefulinmyoptometryclasswork.

[7] S. Rong and B. P. Ng (2021) A deep learning model tocontinuously estimate blood pressure using single channelPPGinProc.IEEEInt.Conf.Acoust.SpeechSignalProcess.(ICASSP), pp. 1200-1204. Minor yet helpful methodology.

[8] A.Gosselin,E.ChovetandV.Lecault(2018)Microfluidicblood typing Advances and clinical perspectives, Lab on aChip, vol. 18, no. 22, pp. 33861400. Micro-fluidics that arecool and might substitute existing kits.

[9] Because contemporary transfusion medicine depends onautomated immunohaematology analysers, this article willexamine theirhistory and the originsof immunohaematologyanalyserscurrentlyutilizedinmedicalpractice.<|human|>T.LRatliffandM.Patel(2019)-Automatedimmunohaematologyanalysers in modern transfusion medicine, TransfusionMedicine Reviews, vol.33, no.4, pp 196-203. Transfusionprotocols my thesis has perfect reference.

[10] U. Raghavendra et al. (2018) -Deep convolution neuralnetworktopreciselydiagnoseglaucomawithdigitalfundus

images, Information Sciences, vol. 441, pp. 41-49.Ridiculously new, and hence my excuse to warrant a projectduring our ophthalmology elective.

[11] S.Kucheryavskiy(2017)-WavelengthselectionofNIR-basedmarijuanaconstituentanalysis,AnalyticalandBioanalyticalChemistry,vol.409pp.34713480.GoodinmyNIRspectrumanalysislaboratory.

[12] R. Padiya and V. Sharma (2017) - Correlation betweenPPGsecondderivativeandbloodviscosityindices,JournalofMedicalEngineeringandTechnology,vol.41,no.6,pp 470-477. Good connection between the theory and measurableparameters.

[13] M. Elgendi (2012), On the analysis of fingertipphotoplethysmogram signals, Current Cardiology Reviews,vol.8,no.1,pp.1425.Thebookisfundamentalreadinginanycardiovascularengineering.

[14] A. Neshitov et al. (2021) - Wavelet analysis and self-similarity of photoplethysmography signals to HRVassessment and emotion recognition since Sensors, vol.21,no.18 p.6136. Embarkation on my psychologyengineeringinterdisciplinary seminar.

[15] R. Kaur and G. Singh (2017) -SVM-based blood groupclassification based on the features of impedanceplethysmography, in Proc. Int. Conf. Computing,Communication and Automation (ICCCA), pp. -812-816.Hypothetical machine-learning example.

[16] Y. Zhang, C. Liu and D. Zhang (2020) Blood groupestimation with photoplethysmography on random forestclassifier, in Proc. IEEE Biomed. Circuits Syst. Conf.(BioCAS), pp.1-4. Concluding my group project with thisinformation.