A Nonlinear Decision Tree based Classification Approach to Predict the Parkinson's disease using Different Feature Sets of Voice Data

Satyabrata Aich[1], Kim younga[2], Kueh Lee Hui[3], Ahmed Abdulhakim Al-Absi[4] and Mangal Sain[5]

[1]Department of Computer Engineering, Inje University, South Korea

[2]Department of Rehabilitation Science, Inje University, South Korea

[3]Dept. of Electrical Engineering, Dong-A University, South Korea

[4]Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea

[5]Department of Computer Engineering, Dongseo University, South Korea

satyabrataaich@gmail.com, kya2664@hanmail.net, leehkueh@dau.ac.kr, absiahmed@kduniv.ac.kr, mangalsain1@gmail.com

Abstract— In the past few years, lot of researchers are working to get some breakthrough for early detection of Parkinson's disease. As the old age population is increasing at a higher rate as well as it is predicted that the old age population will increase to a much higher total by 2050, it's a become a rising concern to the developed countries because the cost due to the healthcare service of these disease is really high. Parkinson's disease (PD) belongs to the group of neurological disorder, which directly affect the brain cells and the effect is shown in terms of movement, voice and other cognitive disabilities. Researchers are keep working on different fields such as gait analysis as well as on speech analysis to find the predictors of the Parkinson's disease. Recently machine learning based approach has been used by many researchers across the field because of its accuracy on the complex data. Machine learning based approach has been used in many cases of Parkinson's disease using gait data as well as voice data. However, so far no body has compared the performance metrics using different feature sets by applying non-linear based classification approach based on the voice data. So in this paper we have proposed a new approach by comparing the performance metrics with different feature sets such as original feature sets as well as Principal component Analysis based feature reduction technique for selecting the feature sets. We have used non-linear based classification approach to compare the performance metrics. We have found an accuracy of 96.83% using random forest classifiers using PCA based feature sets. This analysis will help the clinicians to differentiate the PD group from healthy group based on the voice data.

Keyword— Parkinson's disease, machine learning, feature selection, voice data, performance metrics

Satyabrata Aich is working as a researcher in the field of computer engineering He has over four years of teaching, research and industry experience in India and abroad. He has published many research papers in journals and conferences in the realms of Supply Chain Management and data analytics. His research interests are natural language processing, Machine learning, supply chain management, data mining.
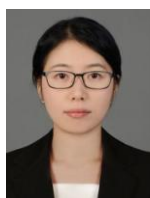
Young a Kim is a licensed Physical therapist and has received her Masters degree in Counseling and Psychology in 2008. She is working as a lecturer for Clinical Kinesiology and Functional Anatomy at Inje University. She is also a PhD candidate in the Department of Rehabilitation Science at the same University. Her research focuses on assessment tools and rehabilitation for the elderly, especially for the detection and treatment of age related conditions. She is a co-author of the book 'Community Occupational Therapy' and a frequent contributor to the journal of 'Physical Therapy Science' and a member in the Korean Ageing Friendly Industry Association as well as the Society of Occupational Therapy for the Aged and Dementia.

Ahmed Abdulhakim Al-Absi is an assistant professor in Department of Computer Engineering (Smart Computing) at Kyungdong University in South Korea. He earned a Ph.D. in computer science from Dongseo University in 2015. He received M.Sc. degree in information technology at University Utara Malaysia in 2011, and B.Sc. degree in computer applications at Bangalore University in 2008. His research interests include Big Data processing, Hadoop, Cloud computing, IoT, Distributed systems, Parallel computing, Bioinformatics, Security, and VANETs.

Kueh Lee Hui is working as an assistant professor at the department of Electrical Engineering, Dong-A University since 2012. She completed her PhD Degrees from Department of Electrical Engineering, Dong-A University, Korea. In 2009 she completed her BS degree in Electronic and Communication, Department of Electronic Engineering, University Malaysia of Sarawak, Malaysia. She also has done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.

Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Information Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 50 international publications including journals and international conferences. He is a member of TIIS and a TPC member of more than ten international conferences.