

A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease

Satyabrata Aich^{*}, Hee-Cheol Kim^{*}, Kim younga^{**}, Kueh Lee Hui^{***}, Ahmed Abdulhakim Al-Absi^{****} and Mangal Sain^{*****}

^{*}Department of Computer Engineering, Inje University, South Korea

^{*}Department of Rehabilitation Science, Inje University, South Korea

^{***}Department of Electrical Engineering, Dong-A University, South Korea

^{****}Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea

^{*****}Department of Computer Engineering, Dongseo University, South Korea

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

Abstract—Among the neurological diseases, parkinson's disease is the second most common disease, which affect the old age people over the age of 65 year. It is also mentioned that the number of people affected with Parkinson's disease will increase at a higher rate until 2050, and it will be a rising concern to many 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. Past few years researchers are working for detection and monitoring of the Parkinson's disease by using the speech analysis as well as the gait analysis data. Machine learning and artificial intelligence techniques are gaining popularity because these techniques are able to automate the pattern recognition process with high accuracy.

However so far, no body has compared the performance metrics using different feature sets by applying nonlinear and linear classification approaches 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 genetic algorithm-based feature sets as well as Principal Component Analysis based feature reduction technique for selecting the feature sets. We have used different classification approaches to compare the performance metrics. We have found an accuracy of 97.57% using SVM with RBF by using genetic algorithm-based feature sets. This analysis will help the clinicians to differentiate the PD group from healthy group based on the voice data.

(Keywords— Parkinson's disease, machine learning, feature selection, voice data, genetic algorithm



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.



Hee-Cheol Kim received his BSc at Department of Mathematics, MSc at Department of Computer Science in SoGang University in Korea, and PhD at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. His primary concerns are u-Healthcare, smart home technology, e-learning and Human Computer Interaction (HCI). He is associate professor at Department of Computer Engineering, Inje University in Korea. He has published many research papers in journals and conferences in the realms of HCI and CSCW.



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.



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 done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.



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.



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 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 THIS and a TPC member of more than ten international conferences.