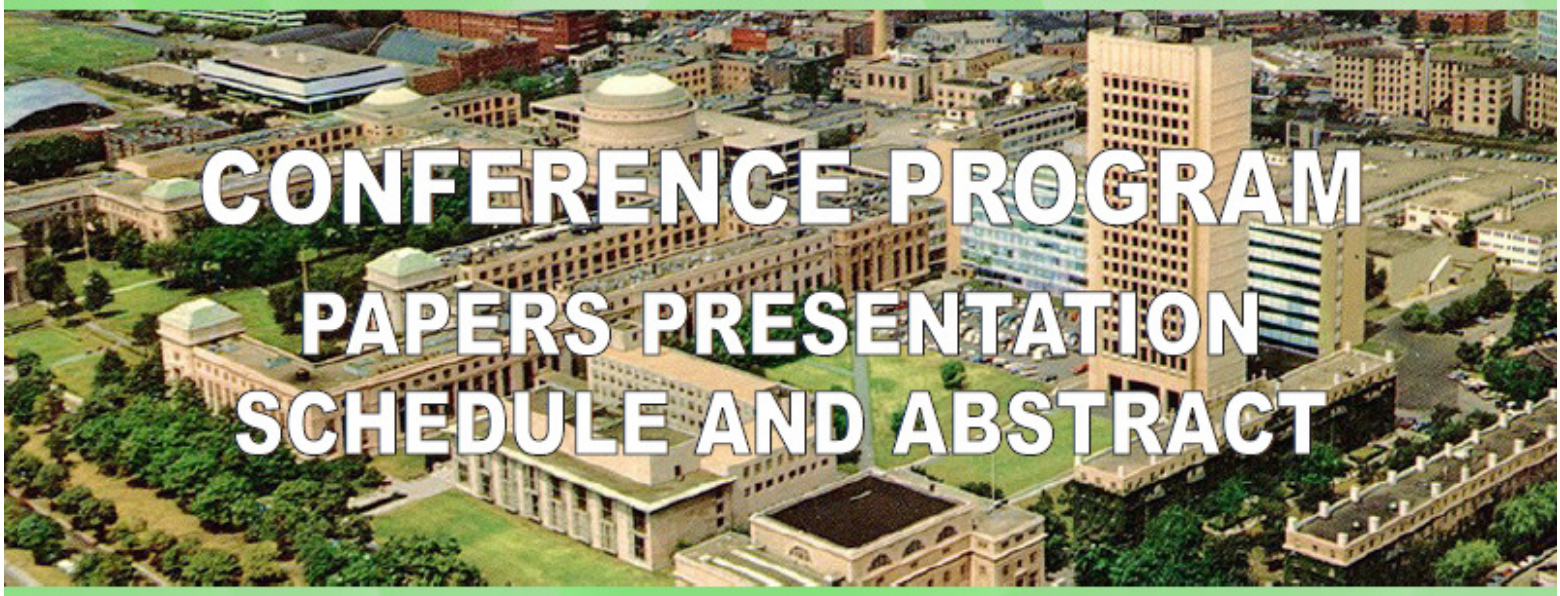
IEEE MIT URTC 2020

UNDERGRADUATE RESEARCH TECHNOLOGY CONFERENCE

October 09 - 11, 2020 | Cambridge, Massachusetts, USA (Virtual)

MEET INNOVATIVE TECHNOLOGY



CONFERENCE PROGRAM
PAPERS PRESENTATION
SCHEDULE AND ABSTRACT

Organized and Sponsored by IEEE Boston Section and MIT IEEE Student Branch

https://urtc.mit.edu

mit ieee
urtc 2020

IEEE Catalog Number: CFP20E50
ISBN: 978-1-7281-7571-3

October 10, 2020 (Saturday)

Technical Paper Oral Presentation (PM Track #1)

EST 12:30pm - 2:30pm HOPIN Session Room A

Machine Learning / Artificial Intelligence (AI) Track #2

Track Chair: James Byleckie

➢ **12:30pm (PA20-0062)**

Time Warping Clustering for the Forecast and Analysis of COVID-19

Qixuan Jin (California Institute of Technology)

This paper presents an effective algorithm for the clustering of confirmed COVID-19 cases at the county-level in the United States. Dynamic time warping and Euclidean distance are examined as the k-means clustering distance metrics. Dynamic time warping can compare time series varying in speed, as counties often experience similar outbreak trends without the timelines matching up exactly. The effect of data preprocessing on clustering was systematically studied. Further analyses demonstrate the immediate value of our clusters for both retrospective interpretation of the pandemic and as informative inputs for case prediction models. We visualize the time progression of COVID- 19 from April 5, 2020 to August 23, 2020. We proposed a Monte-Carlo dropout feedforward neural network with the ability to forecast four weeks into the future. Predictions evaluated from July 24, 2020 to August 20, 2020 demonstrate the better empirical performance of the model when trained on the clusters, in comparison with the model trained on individual counties and the model trained on counties clustered by state.

➢ **12:40pm (PA20-0068)**

Hierarchical BiGraph Neural Network as Recommendation Systems

Dom Huh (George Mason University)

Graph neural networks emerge as a promising modeling method for applications dealing with datasets that are best represented in the graph domain. In specific, developing recommendation systems often require addressing sparse structured data which often lacks the feature richness in either the user and/or item side and requires processing within the correct context for optimal performance. These datasets intuitively can be mapped to and represented as networks or graphs. In this paper, we propose the Hierarchical BiGraph Neural Network (HBGNN), a hierarchical approach of using GNNs as recommendation systems and structuring the user-item features using a bigraph framework. Our experimental results show competitive performance with current recommendation system methods and transferability.

➢ **12:50pm (PA20-0069)**

Positive Unlabeled Gradient Boosting

Andrea Boskovic (Smith College)

Classification applied to medical datasets from diagnosis or survey application data often must address the challenge that such data is weakly labeled, with only some positive labels and the rest of the data instances mostly being unlabeled. Standard classifiers struggle to learn the correct class for these positive but unlabeled instances, particularly within imbalanced datasets. The standard Gradient Boosting Classifier is one algorithm that works well on balanced data with completely labeled examples but performs poorly otherwise. In order to improve upon this state-of-the-art method, we propose a modification to its loss function that empowers it to learn a decision boundary more reflective of the dataâ€™s true distribution. We call this the novel gradient boosting classifier. Our experimental study demonstrates that our proposed new classifier outperforms the state-of-the-art by 8.3% on average across several public medical data sets. This classifier can be applied to healthcare settings, where imbalanced and positive unlabeled data sets are common.

➢ **1:00pm (PA20-0072)**

A Low-Cost Radar-based Domain Adaptive Breast Cancer Screening System

Samuel Claflin (University of Massachusetts Lowell)

Over the past three decades, the advancements of breast cancer screening technologies such mammography, ultrasound, Magnetic resonance imaging (MRI) saved countless lives. The invention of mammography screening of breast cancer in the 1990s led a technological revolution which is, now-a-days, coupled with MRI and/or ultrasound to achieve diagnoses of much greater accuracy than previously attainable. However, these technologies (mammography, ultrasound, MRI) are not as widely available to patients as one might assume less fortunate countries (such as Bangladesh) often cannot afford the potentially enormous price tag that several ultrasound machines of sufficient quality for accurate diagnoses incurs. In this paper, we present a low-cost (<100 USD) millimeter Wave (mmWave) Radar sensor array (3-10 GHz) imaging technology and a deep learning domain adaptation model-based breast cancer screening system. More specifically, (i) we develop a mmWave Radar sensor array (18 sensor antennas) based 2D imaging system; (ii) we develop a deep learning based domain adaptation model that can learn breast segmentation and cancer detection from expensive source data (mammography, ultrasound) and transfer the knowledge to less expensive target data (Radar images), (iii) we validated our system and methods by utilizing our existing mammography and ultrasound breast cancer screening data as well as 14 patients' Radar images collected from a third world country (Bangladesh).

➢ **1:10pm (PA20-0089)**

Utilizing Artificial Intelligence to Diagnose Autism Spectrum Disorder Based on Eye Tracking Saccades

Abinaya Dinesh (Stanford University)

Autism Spectrum Disorder (ASD) refers to a group of neurodevelopmental disorders, usually characterized by impairments in social interaction skills, struggle with verbal and nonverbal communication, and repetitive behaviors. ASD cannot be diagnosed with any one medical test and presents a wide range of symptoms that make it hard to identify in patients of all ages and severities. This paper proposes a novel method of diagnosis to shorten time, resources, and uncertainty during this process. This predictive machine learning model uses eye tracking (ET) and follows saccadic eye movements in patients to identify patterns, irregularities, and fixation time as symptoms of ASD. The results of experimental testing show that the Two Class Boosted Decision Tree is the most accurate model to use, with an estimated accuracy score of 88-91%. As presented by the model, the tracking of sensorimotor activity such as saccades and application into machine learning has the capability to improve the diagnosis of patients across the spectrum.

➢ **1:20pm (PA20-0099)**

Computer-aided Ischemic Stroke Classification from EEG Data Using a Single-Tiered Spiking Neural Network Framework

Elon Litman (John L. Miller Great Neck North High School)

Ischemic stroke is one of the most common cerebrovascular conditions, and constitutes a significant portion of global mortality rates. Early diagnoses are vital for successful recoveries, but with conventional diagnostic imaging techniques and computer systems, radiologists misdiagnose more than 20% of all ischemic strokes. AI methods have been developed for automated neurological disorder prediction using electroencephalographic (EEG) data, but artificial and recurrent classifiers have still seen mediocre performance. Spiking neural networks (SNNs), however, have demonstrated their capacity in personalized spatio-and spectro-temporal non-stationary time series data modeling and pattern recognition. This study developed an SNN model that can detect cerebral ischaemia from temporal EEG data with an accuracy of 94.45%, far exceeding the performance of stroke models operating on MRI data. The approach outlined also required a significantly low number of training samples, having been trained and evaluated on an EEG corpus with recordings from 46 stroke patients and 46 healthy individuals, in addition to being scalable to other neurodegenerative diseases and mental illnesses. Ultimately, the diagnostic precision of the SNN can be adapted in professional environments to replace or be in conjunction with other computerized medical systems to improve ischemic stroke prognosis and recovery.

➢ **1:30pm (PA20-0101)**

Detecting Fake News on Twitter Using Machine Learning Models

Emma Cueva, Grace Ee, Akshat Iyer, Alexandra Pereira, Alexander Rosman (Rutgers University)

As the popularity of social media has risen, people have become increasingly aware of current events, often through sources such as Twitter. One issue with these news sources is the prevalence of false information, or fake news. Even as some social media platforms take initiative with labels or warnings, fake news continues to have dangerous consequences beyond misinformation. The goal of this research is to implement an effective method of identifying fake news spread on Twitter through the use of Artificial Intelligence (AI). More specifically, the investigation studied Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Natural Language Processing (NLP) networks to compare their accuracy when predicting fake news. Data was preprocessed and used to train AI models; figures were then generated for analysis. All three models achieved high accuracy in detecting fake news; however, the NLP model was the only iteration that possessed the ability to identify satire as fake news. For this reason, the NLP model was deemed the preferred choice for detecting fake news on Twitter.

➢ **1:40pm (PA20-0108)**

Determining Top Fully-Connected Layer's Hidden Neuron Count for Transfer Learning, using Knowledge Distillation

Ritwick Ghosh (Indian Institute of Engineering Science and Technology, Shibpur)

Deep convolutional neural network (CNN) assisted classification of images is one of the most discussed topic in recent years. Continuously innovation of neural network architectures is making it more correct and efficient every day. But training a neural network from scratch is a very time-consuming and requires a lot of sophisticated computational equipment and power. So, using some pre-trained neural network as feature extractor for any image classification task or "Transfer Learning" is very popular approach that saves time and computational power for practical use of CNNs. In this paper an efficient way of building full model from any pre-trained model with high accuracy and low memory, is proposed using Knowledge Distillation. Using the distilled knowledge of the last layer of pretrained networks is passes through fully-connected layers with different hidden layers, followed by Softmax layer. The accuracies of student networks are mildly lesser than the whole models, but accuracy of student models clearly indicates the accuracy of the real network. In this way best number of hidden layers for dense layer for that pretrained network with best accuracy and no-over-fitting can be found with less time. Here VGG16 and VGG19 (pre-trained upon "ImageNet" dataset) is tested upon chest X-rays (pneumonia and COVID-19). For finding the best number of hidden layers total it saves nearly 44 minutes for VGG19 and 36 minutes 37 seconds for VGG16 feature extractor.

➢ **1:50pm (PA20-0114)**

A-Seeker: An Efficient Audio Transcription Platform

Harrison Affel, Sean Cox (Wentworth Institute of Technology)

In this digital age, especially during the current global pandemic, consumers are overwhelmed with online audio video content that is produced and distributed at lightning speed. Even though the content is available and accessible, it takes significant time to look for a piece of information needed, resulting in lots of frustration and wasted time. A-Seeker, short for Audio Seeker, is a software platform capable of transcribing and searching audio-video content, enabling users to quickly search for the information they need. The results show that A-Seeker has lower cost and better performance compared to similar solutions.

➢ **2:00pm (PA20-0121)**

A Novel Approach to Toxic Gas Detection using an IoT Device and Deep Neural Networks

Ibrahim Bhavnagarwala, Adam Bhavnagarwala (Danbury High School, NJIT)

Smoking remains one of the top 3 causes of illness in the US, one of top 5 causes of fire hazards in a home and is the single most preventable cause of illness and premature death in the US. The use of Deep Neural Networks in tandem with advances in the sensitivity of gas sensor technology can enable detection of cigarette smoke much sooner and with much higher accuracy than conventional smoke/carbon monoxide detectors used today. The goal of our project was to develop a hardware demonstration and prototype that engages machine learning to not only accurately detect cigarette smoke produced in a room from a single cigarette but also to discriminate cigarettes from other sources of smoke and carbon monoxide such as burning coal, wood or food - typically not possible with conventional smoke detectors.

➢ **2:10pm (PA20-0128)**

Assistive Diagnostic Tool for Brain Tumor Detection using Computer Vision

Sahithi Ankireddy (James B. Conant High School)

Today, over 700,000 people are living with brain tumors in the United States. Brain tumors can spread very quickly to other parts of the brain and the spinal cord unless necessary preventive action is taken. Thus, the survival rate for this disease is less than 40% for both men and women. A conclusive and early diagnosis of a brain tumor could be the difference between life and death for some. However, brain tumor detection and segmentation are tedious and time- consuming processes as it can only be done by radiologists and clinical experts. The use of computer vision techniques, such as Mask R Convolutional Neural Network (Mask R CNN), to detect and segment brain tumors can mitigate the possibility of human error while increasing prediction accuracy rates. The goal of this project is to create an assistive diagnostics tool for brain tumor detection and segmentation. Transfer learning was used with the Mask R CNN, and necessary parameters were accordingly altered, as a starting point. The model was trained with 20 epochs and later tested. The prediction segmentation matched 90% with the ground truth. This suggests that the model was able to perform at a high level. Once the model was finalized, the application running on Flask was created. The application will serve as a tool for medical care professionals. It allows doctors to upload patient brain tumor MRI images in order to receive immediate results on the diagnosis and segmentation for each patient.