CPAMI Founding co-Director and Waterloo AI Institute Faculty Member Prof. Wong Publishes an AI/ML Seminal Article in Nature's Scientific Reports

A milestone paper entitled: "Pattern Discovery and Disentanglement on Relational Dataset" was published in Nature's Scientific Reports on March 11th, 2021, co-authored by Dr. Andrew K.C. Wong, Dr. Peiyaun Zhou, and Dr. Zahid Butt. Dr. Wong, Distinguished Professor Emeritus is the Founding Director in 1984 of the Pattern Analysis and Machine Intelligence (PAMI) research group. He is also a founding co-Director (along with Prof. Fakhri Karray and Prof. Mohamed Kamel of the Electrical and Computer Engineering Department) of the UW Center for Pattern Analysis and Machine Intelligence (cPAMI) in 2011. Dr. Wong is also an affiliate faculty member of the Waterloo AI Institute. Both Dr. Wong and Dr. Zhou are from Systems Design Engineering. Dr. Butt is from School of Public Health and Health Systems, Faculty of Health. WatCo. is giving the team great commercialization support. A clinical application paper, entitled "Explanation and prediction of clinical data with imbalanced class distribution based on pattern discovery and disentanglement", was published in *BMC Medical Informatics and Decision Making* addressing the explanation and prediction ability of thoracic surgical risk.

Pattern Discovery and Disentanglement (PDD) is a novel machine learning (ML) paradigm capable of discovering explicit association patterns from relational data with various sizes, imbalanced groups, and anomalies. The patterns are succinct, comprehensive, concise, and easily tracked and understood. It improves prediction accuracy and facilitates transparent interpretation of discovered patterns in an explicit *Knowledge Base, which* links a) the sources (mathematic spaces grouping specific associations, e.g. association relating to underlying factors that cause patients' symptomatic manifestation); b) the association patterns (e.g. statistically significant health profile and signs-and-symptoms of patients) and; c) the cohorts (e.g. individuals or groups of patients in summarized and comprehensive forms). Hence, PDD promises broad and in-depth applications in genomics, bioinformatics, clinical practices specifically, and ML at large.

The significance of PDD lies in its novel and unique theoretic concept, its algorithmic efficacy, its all-in-one integrated deep knowledge discovery and representation process. It is a game-changer in ML as it presents a breakthrough in the very difficult problem of pattern entanglement that complicates the ML search process and the generation of useful results. It uses the deep knowledge it discovers to rectify the ground truth, direct construction and execution of the classifier/predictors, and improves the predictive accuracy by removing biases. It has also overcome crucial problems on anomalies, outliers, biases, rare and imbalanced groups/classes, and lack of transparency that have been confronting existing ML models.
Dr. Wong and his co-authors have published a number of other important articles in the field.

The real-world problems PDD has solved include:

a) Discovering pattern clusters from protein receptors related to gene classes and functional domains
b) Detecting outliers and anomalies from histopathological data of breast cancer and microarray data of colon cancers and improving prediction accuracy ~10% after the identification and removal of biases and misinformation.
c) Revealing rare patterns for Peritoneal Dialysis Eligibility decision
d) Explaining and predicting thoracic surgery risk from data with imbalanced class distribution
e) Discovering frequent and rare patterns for cannabis use among adolescents in Canada

Since the results PDD obtains are robust, explicit, displayable and explainable for experts' interpretation, interactive querying and knowledge base construction, it has great potential to render a new form of Explainable AI. It bridges the 'AI chasm'— the gap between creating a scientifically sound algorithm and its application to real-world problems. It will play an important role in empirical data sciences as it brings AI closer to domain experts with insights and accountability, meeting the scientific, economic, legal and social challenges for AI in healthcare and data analytics for the years to come.