# CALIFORNIA STATE SCIENCE FAIR
# 2014 PROJECT SUMMARY

**Name(s)**

**Yousuf M. Soliman**

**Project Number**

34654

**Project Title**

## Indium: Using Novel Machine Learning Algorithms to Develop a Nondisease-specific Personalized Medicine Engine

**Abstract**

**Objectives/Goals**

In medicine today, treatments for patients are based on the diagnosis, not the patient individually. It has been shown that this is only effective for about 60% of people. To help treat more patients I developed Indium, a nondisease-specific personalized medicine using novel machine learning algorithms. I divide this problem into 3 distinct steps: diagnosis, prognosis and individualized treatment creation.

**Methods/Materials**

To truly create a completely generalized diagnostic system I developed a powerful natural language processing (NLP) engine that analyzes how language acts on itself. I then connect this NLP to PubMed, a database of medical research papers, to extract features and biomarkers that are indicative of certain diseases. Analyzing the effectiveness of my system, I have found that my generalized algorithm is more effective, by about 3 standard deviations, than the state of the art techniques and physicians. My prognostic software works by utilizing fuzzy lagged data co-clustering, an NP-complete problem. To circumvent this problem I developed a Monte-Carlo approximation that runs in polynomial time. Lastly, dealing with treatment creation I developed a Q-learning algorithm that dynamically adjusts for the specific patient parameters. To deal with the problem of censored data, I created an SVM system to maintain a constant belief state of the subject. I demonstrated the performance of the proposed algorithmic framework through the analysis of real clinical trials.

**Results**

The personalized medicine system I developed not only operates in non-optimal environments, but it is more effective than the state of the art techniques. Since my algorithm is offloaded into the cloud I am able to help patients around the world regardless of their socioeconomic status.

**Conclusions/Discussion**

If one can predict the likely result of a sequence of actions or treatments for some time out into the future, then they can use that to determine the optimal action right now.

The work presented here adds to a growing body of evidence that such complex treatment decisions may be better handled through modeling than intuition alone. This is true due to the fact that it has been shown to more accurately predict the optimal treatment plan than trained physicians.

**Summary Statement**

I developed powerful artificial intelligence algorithms to determine the optimal treatment path for specific patient by analyzing the entire patient history at an individual level.

**Help Received**

I consulted with several professors and physicians about the eminent need for personalized medicine.