Advanced Knowledge Enablement (CAKE)

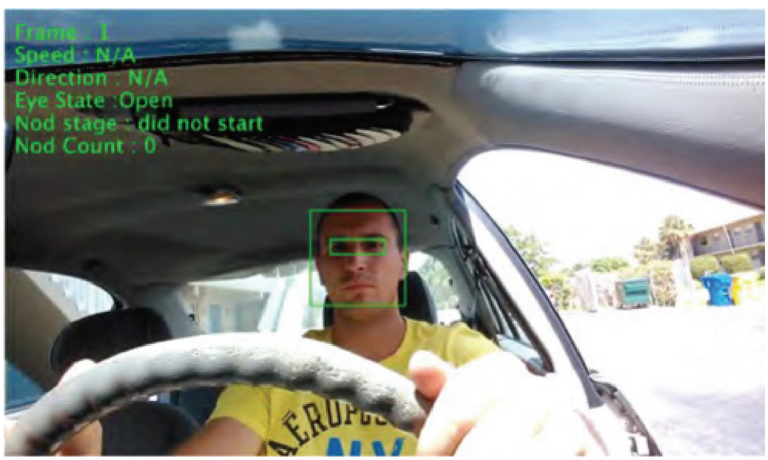
Florida International University, Naphtali Rishe, 305.348.2025, rishen@cs.fiu.edu
Florida Atlantic University, Borko Furht, 561.297.3180, borko@cse.fau.edu
Dubna International University, Evgenia Cheremisina, 7.49621.22.683, e.cheremisina@geosys.ru

Driver's Drowsiness Detection System

In this project, CAKE researchers at Florida Atlantic University developed a driver drowsiness detection system. This research combines smartphones with machine learning to detect driver drowsiness in real time. The system is based on visual input that includes the position of a driver's face and head/eye position.

This innovative algorithm combines software components for face detection, human skin color detection, and a classification algorithm for the eye state (open vs. closed). The system uses commercially available smart phones to monitor drivers. Visual inputs are then used to detect signs of drowsiness. When signs of drowsiness are detected the system issues an alert.

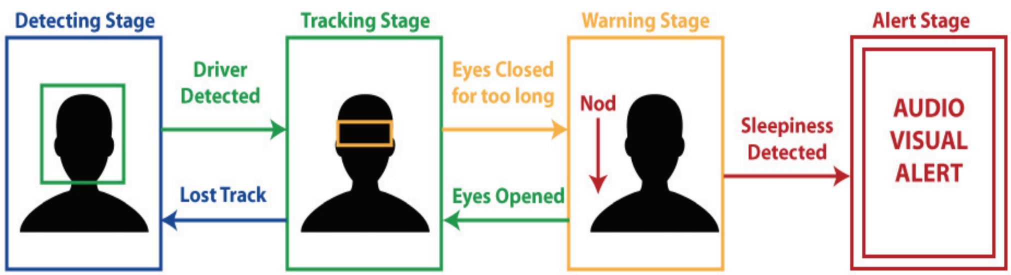
The system uses innovative machine learning algorithms that continuously monitor driver behavior and alerts the driver in real time when certain thresholds are met. The high speed algorithms provide continuous, real-time analyses of driver imagery without consuming an undue of battery power.



In order to capture a clear view of the driver's face the device is mounted securely and aimed precisely at a specific distance. Once mounted, the device begins a training process to determine the baseline head position and eye movement. After this calibration, real time monitoring begins, sending an alarm if the head dips or eyes close for prolonged periods; both can be indicators that the driver is falling asleep. The system has been demonstrated to be applicable in real time, easily portable to different platforms, highly accurate, and robust.

The advantage of being able to use the ubiquity of smartphones instead of relying on built-in products makes it feasible to deploy the system in any vehicle. With today's inexpensive infra-red cameras the device can operate in poor lighting conditions.

The CAKE team has developed and tested the software prototype for Android smartphones. The next step is to port to iOS platform for iPhones. Soon the system will be commercialized as a general smartphone application. Soon an embedded system with an infra-red camera will be completed. It will use the same algorithms so it can be mounted on the car's dashboard. These two systems, one using smartphones and another using a device, will be universally-applicable to all cars, and at very low cost. Eventually, the systems can be built in into newly manufactured cars.



The four stages of the Drowsiness Detection System.

With the high rates of accidents caused by drowsy drivers the need for these products is obvious. Beyond passenger vehicles, drowsiness alert systems can be expected to be applicable to public transportation modalities because taxies, buses, trains, subways, and long-haul truckers all face the same risks from driving drowsy. Adoption in these sectors could be driven by regulatory agencies as part of safety requirements.

> **Economic impact**: There is significant statistical evidence that there is a commercial need for driver drowsiness detectors. There are alarming numbers of road accidents. Worldwide, over a million people die on the road every year with 6% being linked to driver drowsiness. It has been estimated that nearly 75,000 deaths could be avoided by alerting drivers, either startling them awake, or indicating that they should pull over and sleep instead of continuing to endanger themselves or other on the road. Though almost impossible to estimate precisely, the economic impact of driver drowsiness detection algorithms and related products substantial.

For more information, contact Borko Furht at Florida Atlantic University, bfurht@fau.edu, Bio http://www.cse.fau.edu/~borko/, 561.297.3180.

Modeling Ebola Spread Using Big Data Analytics

A model of Ebola spread was developed using innovative big data analytics techniques and tools. Massive amounts of data were used from various sources including Twitter feeds, Facebook and Google. This data was then fed into a decision support system. It models the spread pattern of the Ebola virus and creates dynamic graphs and predictive diffusion models based on the outcome and impact on either a specific person or on entire communities.

This CAKE research created computational spread models for Ebola that will potentially lead to more precise forward predictions of disease propagation. The tool will help identify individuals who are possibly already infected. It is capable of performing trace-back analysis to locate the possible sources of infection for particular social groups. Working with Florida International University researchers and other partner universities, the Florida Atlantic University research team also collaborated with LexisNexis Risk Solutions (LN); a leading big data company and CAKE member. LN provided large amounts of data about relationships among people in the United States. Researchers used data analytics and tools to model disease spread patterns. The project performed modeling, analytics, and development of a Decision Support System (DSS), that calculates probabilistic outcomes of Ebola impact on either specific persons or communities at a specific locations. This information is then fed to FIU's Terafly system for geospatial mapping and other services.

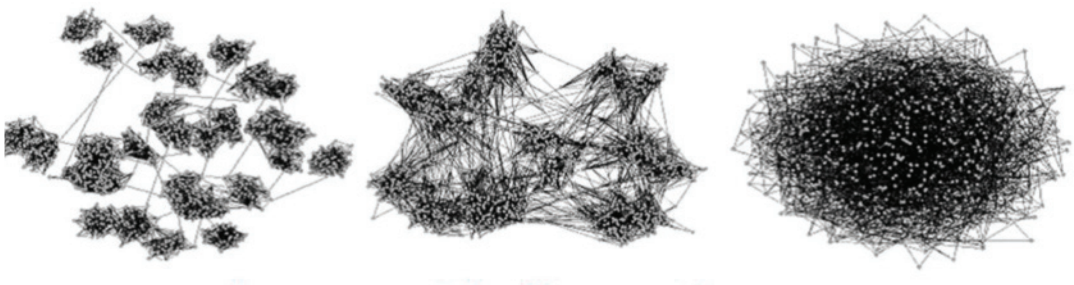
Jointly with the LN research team, people clusters were created based on proximity and a model using weighted scores was built to approximate physical contacts. In creating people clusters, researchers used public record graphs to calculate distances between an affected person and his/her relatives and friends. Based on this model, the disease propagation paths were developed.

As part of this research, the NewsCubeSum was developed. This is a personalized multidimensional news update summarization system. It collects data from news articles and Twitter. The system utilizes OLAP (OnLine Analytical Processing) and supervised sentence selection techniques to generate brief summaries. It delivers news summaries in multiple dimensions, such as time, entity, and topic. This project demonstrated how this system can be used in improving situational awareness during disease outbreaks.

Tracking and containing Ebola requires enormous resources. This system provides a proactive approach to reasonably reduce the risk of exposure of Ebola spread within a community or a geographic location. This work represents an improvement over previous state-of-the-art, because it used innovative data analytics techniques and the latest HPCC Systems technology to developed models of Ebola spread. With information from multiple sources indicating infected individuals and their personal relationships and social groups, dynamic graphs can be created, and predictive diffusion models can be used to study key issues of Ebola epidemics, e.g., location, time and number of expected new cases. The two fundamental diffusion models are Independent Cascade Model (IC) and Linear Threshold Model (LT). Both models follow an iterative diffusion process wherein infected nodes infect their uninfected neighbors with certain predictable probabilities. Based on fundamental models, advanced propagation models were developed to estimate an influence function by examining past and newly infected notes and predict subsequent infections.

The DSS allows the individual to enter their specific information through a Web-application such as travel information. The mobile interface application would automatically extract the geo-coordinates of the individuals. This would allow the system to intelligently query the movement of the person and a possible contact in the areas affected by Ebola or in the areas affected by Ebola. For instance, if found that a person affected with Ebola was in a theater the previous evening, then a monitoring alert could be issued to the other people within the theater to take extract precaution and watch any sign of Ebola impact such as fever or headache.

The mobile interface will further allow the people to enter the signs of Ebola such as fever specifically above 100.4^{o}, chills, headache and vomiting, myalgia, intense weakness. While these signs are very common with other diseases as well, such as Malaria and Typhoid, laying over these impacts on the geo-coordinates of having the person in a geographic proximity of a person in a community with impact of Ebola, will provide more focused results with higher accuracy.



Social network analysis types showing increasing risks of Ebola propagation

The DSS will also interface with the social groups of the person. This allows the extraction of relationship maps to predict the higher probability of Ebola spread within specific communities. The social group integration such as a LinkedIn or a Facebook application programmable interface (API) can extract such data very accurately thereby providing high precision and accuracy in risk prediction.

> **Economic impact**: The proposed methodology, including coalition-building efforts, supports solutions to a wide range of other public health issues. The economic impacts can be tremendous by predicting outbreaks of the deadly Ebola virus (or any other epidemics) and directing potential victims to the nearest suitable medical facility. This research can also help in indicating the areas in which new facilities should be opened, where disease outbreaks are beginning to occur, and how they are likely to expend.

For more information, contact Borko Furht at Florida Atlantic University, bfurht@fau.edu, Bio http://www.cse.fau.edu/~borko/, phone 561.297.3180.

Smart Building Optimization Systems

The ongoing development of smart buildings has recently gained in importance at Florida Atlantic University. Innovative methods are being used to optimize the operation of buildings and thereby reduce operating expenses and reduce energy consumption. In this project, research was conducted on the FAU's new LEED Platinum-certified Engineering building that houses the Center for Advanced Knowledge Enablement's (CAKE's) laboratories. The building was designed both as a model of how new technologies can drastically decrease the energy requirements of a large building and as a "living laboratory," so that students and faculty can actually see how these systems work and interrelate. The building is equipped with hundreds of sensors that measure and collect various parameters and display them in real-time on a dashboard, which is accessible through a Web-based application called DeviseWise from ILS Technology, a member of our I/UCRC and our partner in this project.

Mechanical equipment, pumps, piping, and sensors are part of the building HVAC systems.

During 2014-15 the CAKE team developed a data warehouse that stores information from several different sensor systems including DeviceWise, standalone wireless and wired sensors, PDM calculated clusters, and weather stations. The collected weather data comes from a link to the WeatherBug API. Every 15 minutes it pulls meteorological information including temperature, humidity, wind speed and direction, air pressure, rain amount and light levels. The system provides tools for extracting and analyzing sensor data. It exports it in a variety of formats for use by other tools such as the Weka Machine Learning Suite and Excel.

As part of the NSF Center project, FAI faculty and students worked with Aware Technologies and their Process Data Monitor (PDM) system. This is an alerting system that uses data mining techniques to categorize sensor data into similar clusters of information. It automatically detects when a current cluster is outside of normal operating parameters. It then reports anomalies to an operator.

Researchers have analyzed the efficiency of the building's solar panels by tracking the power generated over a one-year period. They noticed high variations in the amount of solar power generated. This was mainly attributable to variable day-by-day cloud coverage. Between March and May 2012 there were very clear skies with almost no rain and hardly any clouds. Conversely, the beginning of 2013 witnessed a period of high cloud coverage combined with shorter days. This caused reduced solar energy generation.

Currently, the building's solar energy production is on average 4.45% of the total energy used by building systems.

Solar photovoltaic cells on a building roof for power generation.

Advantages over previous methodologies: This work presents an excellent example how researchers are using the new Platinum-certified building and FAU labs as a "living laboratory" in order to better understand and conduct research in the areas of smart building technologies and optimization. The main focus is on improving energy efficiency; however the other components of the project include sustainability, water savings, material and resource selection, and indoor environmental quality.

The end-user systems can be summarized as: 1) Alerting and monitoring systems that analyze data from various sensors in the building to issue alerts when anomalies are detected, and; 2) A data warehouse system that collects and stores data from the sensor systems. Data can then be used for determining correlations between photovoltaic energy generation and weather conditions, and calculation of energy flow between the different components of the air conditioning systems.

> **Economic impact**: Strong instrumentation in the new LEED Platinum engineering building opens up a multidimensional view of the inner working of its HVAC and power systems. A variety of sensors allow detailed analyses of the building system performance and the data center power consumption. This research provides the foundation for analyzing and improving energy efficiency in buildings equipped with sensors and solar panels. The results of this research are being used to analyze and improve the operation of new "smart" buildings.

For more information, contact Ionut Cardei at Florida Atlantic University, icardei@fau.edu, Bio http://www.cse.fau.edu/~icardei/index.html, 561.297.3401.

Machine Learning Algorithms on HPCC/ECL Platform

The FAU's Center for Advanced Knowledge Enablement (CAKE) developed a wide range of machine learning algorithms on the High Performance Cluster Computing (HPCC) platform. This platform includes traditional algorithms such as Naïve Bayes and K-Nearest Neighbors, to more advanced techniques such as Deep Learning. This enables researchers and practitioners to apply machine learning algorithms on big data to extract patterns and perform predictive tasks. The HPCC architecture, written in conjunction with the ECL programming language, is LexisNexis's answer to applying machine learning methods on big data.

HPCC provides a platform for implementing parallel, distributed, and scalable machine learning algorithms. The general linear algebra and statistical operations implemented in HPCC along with the data structures provide an ideal platform for implementing the machine learning algorithms.

Machine learning algorithms using the vastly increasing volumes of personal, wearable health information to anticipate health risks and concerns given the user's historical and present lifestyle information.

The amount of data produced from bioinformatics to social media to web documents, has exploded in recent years. Many traditional approaches fail when dealing with terabyte, petabyte, or larger datasets preventing consumers from fully benefiting from their data. Analyzing these large amounts of so-called big data opens up new research areas that were not possible 20 years ago. Major companies such as Google, Facebook, Twitter, and Amazon would be less effective without taking advantage of big data analytics. Big data analysis demands large-scale systems to both manage and process huge quantities of data. Big data with characteristics such as high volume, high complexity, and data heterogeneity require new ways of thinking and new paradigms for knowledge extraction.

This breakthrough is an improvement over more traditional analytics which rely heavily on human analysts. The sheer volumes of data that have high amounts of potential correlations and hidden patterns do not allow for comprehensive analysis using traditional data analytic tools. Machine learning and big data tools overcome these problems by leveraging properties intrinsic to the data to infer semantics and formats, deriving effective and general algorithms for data processing and analytics.

In comparison to traditional analytics, machine learning delivers the promise of extracting patterns in an automated fashion with far less reliance on human supervision. The methods are data driven. They thrive on and benefit from increased data because with more information, more can be learned. Given the limitations of human comprehension in the face of truly massive amounts of data, machine learning is able to discover hidden patterns on very large-scales.

End users/consumers receive multiple advantages from applying machine learning using big data to do many different kinds of predictive tasks. These are in various domains ranging from fraud detection and product recommendations to energy load forecasting to healthcare predictions. Example end-user products include: 1) Machine learning algorithms using the vastly increasing volumes of personal, wearable health information, via devices such as FitBit, to anticipate health risks and concerns given the user's historical and present lifestyle information; 2) HPCC big data processing and machine learning methods to analyze crime patterns to anticipate specific areas likely to have near term criminal activities and to adjust police resources accordingly; 3) Frameworks to collect and analyze historical and real-time power plant information to predict site-wide and individual component failures producing warnings and predictive work orders mitigating costs due to failures and equipment replacement, and; Analytics based on energy smart grid network data to optimize electrical loads and to anticipate specific failure points in order to avoid loss of service by fixing issues remotely or by dispatching crews prior to anticipated failures.

> **Economic impact**: Machine learning provides reliable and accurate predictions for a wide range of domains. Fraud detection and prevention has contributed to billions of dollars in recovered funds. In 2013, the IRS prevented or recovered $24.2 billion related to identity theft fraud. Predicting shopping patterns can allow stores to better anticipate customer needs and increase overall sales revenue; thus shifting the customer interaction paradigm from a mostly reactive process into a more proactive one wherein systems predict consumer needs and offer optimized services and products. Predicting diseases can reduce burdens on healthcare. Cyber security is critical for national security. Big data and machine learning approaches work synergistically to anticipate and prevent costly attacks. The HPCC architecture provides an established compute resource where big data and machine learning go hand-in-hand without new or additional development and storage costs. The use of ECL to implement the machine learning algorithms allows for better. Since the ECL language is automatically compiled into C++, a more traditional imperative/procedural lower level language, it's easier to measure efficiency gains by comparing the number of lines of code between those lines of code manually written in ECL versus lines of C++ code that are automatically generated by the same ECL codebase. This ratio is roughly 100/1, indicating two orders of magnitude improvements in coding efficiency by leveraging the higher level declarative dataflow paradigm that ECL offers.

For more information, contact Taghi Khoshgoftaar at Florida Atlantic University, khoshgof@fau.edu, Bio www.cse.fau.edu/~taghi/, 561.297.3994.

Modeling Sea Surge and Flooding using ALTA and TerraFly

The TerraFly team at the NSF I/UCRC CAKE at Florida International University, CAKE's member ALTA (Autonomous Lighter Than Air) Systems, Inc., and the SeaRobotics Corporation are producing smart balloons tethered to small unmanned vessels to collect environmental data. In combination with the geospatial data already served by TerraFly, continuing work on this breakthrough has the potential to transform the modeling of sea surges and flooding.

TerraFly (http://TerraFly.com) is a technology and tool set for fusion, visualization and querying of geospatial data. The visualization component of the system provides users with the experience of virtual "flight" over maps comprised of aerial and satellite imagery overlaid with geo-referenced data. Autonomous Lighter Than Air was invented by John Ciampa, the inventor and founder of Pictometry. The ALTA invention has been awarded three U.S. Patents. ALTA has sponsored research at FIU at about $1 million. SeaRobotics Corporation, located in Stuart, Florida, specializes in marine robotics. SeaRobotics Unmanned Surface Vehicles (USVs) are used worldwide by government organizations, academia, commercial survey companies, and others.

TerALTA balloon oblique multispectral aerial image, vegetation index in reddish color, derived nearshore sea depths and isobaths for NSF IUCRC Compendium CAKE-MSS Modeling Sea Surge and Flooding using ALTA and TerraFly

This work represents an improvement over previous state-of-the-art because the addition of an aerial imaging and communication source provides valuable new sensing capabilities. The collection of sea depths near the shore has eluded traditional collection platforms such as LIDAR and aircraft-based aerial photography. The unmanned shallow draft vessel and the low altitude balloon aerial platform are uniquely suited to this task. The TerraFly system allows users to fuse and explore multi-source geospatial data. Examples of fusing ALTA oblique imagery with core geospatial data are at http://Teralta.com.

TerALTA balloon-mounted bathymetry sensor operated via a robotic vessel for NSF IUCRC Compendium CAKE-MSS Modeling Sea Surge and Flooding using ALTA and TerraFly

The project is expected to result in a packaged product that would enable users to collect data in areas that are presently hard to access and to then immediately analyze them via the TerraFly system. Service packages capable of data collection, storage, and analysis are also possible for end-users who may then desire to periodically collect data or who have interest in particular locations at particular times.

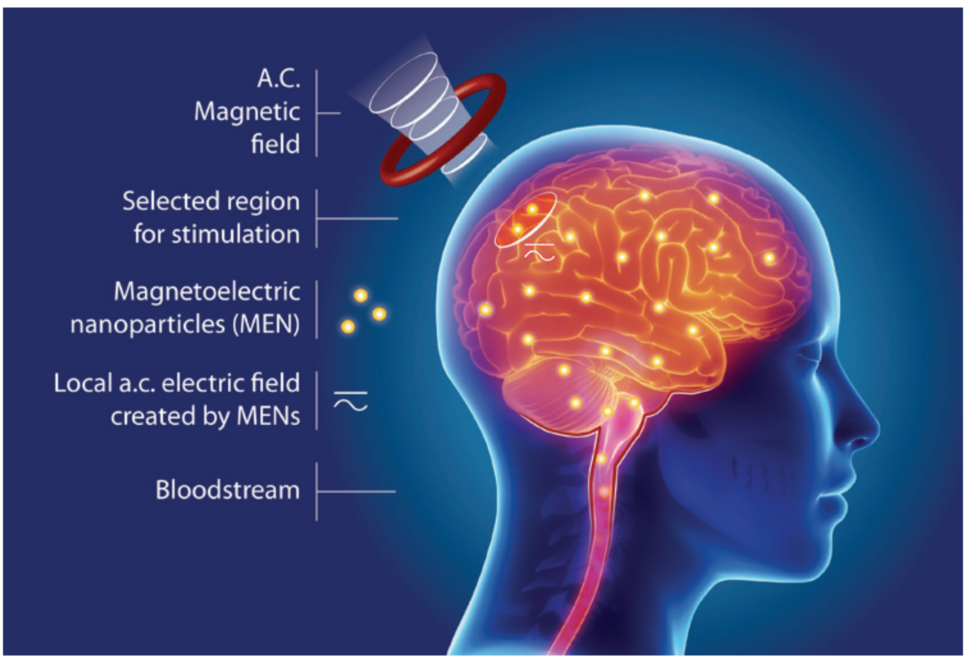
Economic impact: With the help of FIU's TerraFly technology, ALTA opens a new multi-billion market for aerial photography. This is because it produces images of much higher resolution than are possible using other technologies. Image collection is currently accomplished from ground-based cameras, aircraft, and satellites. ALTA is higher than a ship-based camera and therefore sees more. ALTA is lower than an aircraft and therefore sees better. Additionally, compared to other aerial platforms ALTA has low cost components. For the capital outlay of one manned aircraft, hundreds of ALTAs can be deployed. Additionally, balloons are dramatically cheaper than drones. ALTA missions eliminate cost of pilots, aircraft, and airports. Costs of operation are a fraction of those of other aerial collection platforms. Moreover, ALTA can be deployed in minutes and have information and images returned instantaneously. The ALTA technology is thus poised to produce much higher-quality imagery at much lower cost than current technologies, thus opening up new markets and bringing new capabilities to existing markets such as public safety, real estate, construction, environmental monitoring, disaster mitigation and disaster recovery.

For more information, contact Naphtali Rishe at Florida International University, rishen@cs.fiu.edu, Bio http://cake.fiu.edu/Rishe, 305.348.1706.

Wireless Deep-Brain Stimulation With Magnetoelectric Nanoparticles

The brain is a complex bio-electric circuit made of billions of neurons that are inter-connected through chemical and electrical synapses. The ability to remotely stimulate selective neurons deep in the brain remains a major challenge. Overcoming it will enable highly personalized "pin-point" treatments for neuro-degenerative diseases such as Parkinson's and Alzheimer's Diseases, Essential Tremor (ET), Epilepsy, and others. Furthermore, by the law of reciprocity, this nanotechnology can pave a way for reverse-brain engineering.

This FIU team has invented and patented a technology (S. Khizroev and M. Nair, "Wireless brain stimulation," U.S. Patent application 13/900,305, filed 05/22/2013, granted 01/26/2016) to answer the above challenge by using a novel class of multifunctional nanoparticles known as magnetoelectric nanoparticles (MENs). Because of MENs capability to couple magnetic and electric fields at the sub-neuronal level, they enable a unique way to combine the advantages of both the high efficacy stimulation by the electric fields and the external-control capability of the magnetic fields. They therefore open a novel pathway to control the brain.



Wireless deep-brain stimulation with Magnetoelectric Nanoparticles administrated into the brain via an IV injection.

This study, conducted on mice, demonstrated for the first time the feasibility of using MENs as externally controlled "smart" nanoparticles for wireless navigation and selective control of specific functions deep in the brain. The paper recently published by FIU investigators has been selected in the list of 100 Top Science and Technology Stories of the Year (2016) by *Discover* Magazine (January 2016 issue) [R. Guduru, P. Liang, J. Hong, A. Rodzinski, A. Hadjikhani, J. Horstmyer, E. Levister, and S. Khizroev, "Magnetoelectric spin on stimulating the brain," *Nanomedicine* (London) 10 (13), 2051-2061 (2015)]. On behalf of FIU, the team has put together several multi-million-dollar research proposals on the subject.

This breakthrough study represents an important milestone in deep-brain stimulation because it provides a wireless and non-invasive way to achieve significant results. Current DBS technology is operated at macro-scale and often relies on highly-invasive direct-contact-electrode techniques. Current non-invasive brain stimulation methods include rTMS and tDCS, but in both the depth and locality of focus are strongly limited. FIU's technology, using MENs, overcomes the current technology's roadblocks.

The completed study is a stepping stone towards the development of a precision nanotechnology for simultaneously achieving the following three important functions: 1) stimulation; 2) release of drug(s) and other macromolecule(s), e.g., peptides, RNAs, and others, in selective brain regions via remote control, and; 3) mapping the electric field due to neural activity. Achieving each of these functions would be important milestones on their own. Achieving all these three functions simultaneously may very well present a pathway to next-generation pinpoint treatment of neurological diseases.

In the near-term future, a main end-user product would be an advanced wireless deep-brain stimulation (DBS) technology for treating patients with Parkinson's Disease, Essential Tremor, Alzheimer's Disease, Autism, and other neurological diseases.

Even more far-reaching applications might be envisioned when biodegradable MENs will be developed in the future. This is definitely within reach due to the recent development in the emerging field of carbon based nanotechnology. Potential applications span from the prevention and treatment of neurodegenerative disorders to opening pathways to significantly improving fundamental understandings of the brain and to reverse-engineering the brain.

> **Economic impact**: According to the 2011 technology assessment by experts at California Institute of Technology, the deep-brain stimulation (DBS) market was around $360 million. Conservative projections indicate that the technology can impact the multi-billion-dollar medical and information processing market segments.

For more information, contact PI Sakhrat Khizroev at Florida International University, khizroev@fiu.edu, Bio https://ece.fiu.edu/sakhrat-khizroev/, 305.348.3724, or Naphtali Rishe, rishe@fiu.edu, Bio http://cake.fiu.edu, 305.348.1706.