DISCLOSURES

• None
OUTLINE

• Overview of medical device regulatory framework
  – Regulatory overview
  – ML-based medical devices
  – Software as a medical device (SaMD)

• Assessment
  – Imaging-based machine learning (ML) SaMD assessment
  – Framework for assessing ML SaMD modifications
Protect and promote the health of the public by ensuring the safety and effectiveness of medical devices and the safety of radiation-emitting electronic products.
## Device Class & Pre-Market Requirements

<table>
<thead>
<tr>
<th>Device Class</th>
<th>Controls</th>
<th>Premarket Review Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>General Controls</td>
<td>Most are exempt</td>
</tr>
<tr>
<td>(lowest risk)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>General Controls, Special Controls</td>
<td>Premarket Notification [510(k)]</td>
</tr>
<tr>
<td>Class III</td>
<td>General Controls, Premarket Approval</td>
<td>Premarket Approval [PMA]</td>
</tr>
<tr>
<td>(highest risk)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Device Class & Pre-Market Requirements

<table>
<thead>
<tr>
<th>Device Class</th>
<th>General Controls</th>
<th>Premarket Review Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Most are exempt</td>
<td>Demonstrate substantial</td>
</tr>
<tr>
<td>(lowest risk)</td>
<td></td>
<td>equivalence to predicate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>device</td>
</tr>
<tr>
<td>Class II</td>
<td>General Controls</td>
<td>Premarket Notification</td>
</tr>
<tr>
<td></td>
<td>Special Controls</td>
<td>[510(k)]</td>
</tr>
<tr>
<td>Class III</td>
<td>General Controls</td>
<td>Premarket Approval</td>
</tr>
<tr>
<td>(highest risk)</td>
<td>Premarket</td>
<td>[PMA]</td>
</tr>
<tr>
<td></td>
<td>Approval</td>
<td></td>
</tr>
</tbody>
</table>
**Device Class & Pre-Market Requirements**

<table>
<thead>
<tr>
<th>Device Class</th>
<th>Premarket Review Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I (lowest risk)</td>
<td>Most are exempt</td>
</tr>
<tr>
<td>Class II</td>
<td>De Novo</td>
</tr>
<tr>
<td>Class III (highest risk)</td>
<td>Premarket Approval [PMA]</td>
</tr>
</tbody>
</table>

- **Means for new device, without a valid predicate, to be classified into Class I or II**
## Device Class & Pre-Market Requirements

<table>
<thead>
<tr>
<th>Device Class</th>
<th>Controls</th>
<th>Premarket Review Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I (lowest risk)</td>
<td>General Controls</td>
<td>Most are exempt</td>
</tr>
<tr>
<td></td>
<td><strong>Demonstrate reasonable assurance of safety and effectiveness</strong></td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>Special Controls</td>
<td>Premarket Notification [510(k)]</td>
</tr>
<tr>
<td>Class III (highest risk)</td>
<td>General Controls Premarket Approval</td>
<td>Premarket Approval [PMA]</td>
</tr>
</tbody>
</table>
EXAMPLES OF ML-BASED MEDICAL SOFTWARE

FDA News Release

FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients

February 13, 2018

Viz.Ai

FDA News Release

FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

April 11 2018

IDx-DR
ML-BASED MEDICAL DEVICES ARE NOT NEW

- Mostly imaging or physiological signal analysis applications
  - ECG signal analysis
  - Analysis of radiology images
  - Analysis of cytology/pathology images

- Semi-automated cervical cytology slide reader
- Reduce false-negatives due to human error
- FDA approval in 1994
ML-BASED MEDICAL DEVICES

Potential to fundamentally transform the delivery of health care:

E.g., Earlier disease detection, more accurate diagnosis, new insights into human physiology, personalized diagnostics and therapeutics

Ability for ML to learn from the wealth of real-world data and improve performance

Already seeing ML lead to the development of novel medical devices
ML-BASED MEDICAL DEVICES: CHALLENGES

• Need for large, high quality, well-curated data sets
• Explainability of “black box” approaches
• Identifying and removing bias
• Oversight to ML-based algorithms that learn/change over time
“Software as a Medical Device” (SaMD) is defined as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.

5http://www.imdrf.org/workitems/wi-samd.asp
# SaMD Risk Categorization

### Increasing Significance

<table>
<thead>
<tr>
<th>State of Healthcare Situation or Condition</th>
<th>Significance of Information Provided by SaMD to Healthcare Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treat or Diagnose</td>
</tr>
<tr>
<td>Critical</td>
<td>IV</td>
</tr>
<tr>
<td>Serious</td>
<td>III</td>
</tr>
<tr>
<td>Non-Serious</td>
<td>II</td>
</tr>
</tbody>
</table>


5/9/2019
FUNDAMENTALS OF IMAGE-BASED ML ASSESSMENT

• Device description
• Data
• Performance assessment
  – Standalone performance
  – Reader performance (when appropriate)
  – ...
• Human factors or other information/testing as appropriate
• ...

5/9/2019
DEVICE DESCRIPTION

• Device & algorithm descriptions
  – Device usage (mode of operation, patient population, …)
  
  – Algorithm design and function
    • Including structure of traditional and deep learning networks
    • Inputs
      – Type and range of signals/data
    • Outputs

  – Training process

  – Training/test database
  – Reference standard
  – …
ML algorithms are data-driven
  - Versus, for example, physics or biology based
ML algorithm development now facilitated by standardized ML platforms
  - Brings ML to a wider array of users
  
  - The good
    • Access to high-quality data streamlines design of novel ML applications
  
  - The bad
    • Garbage in - garbage out
Performance Testing

- Performance of ML algorithm on an independent data
  - Ideally, identifies problems with training process
PERFORMANCE TESTING

• Standalone performance
  – Performance of algorithm alone
  – Assesses robustness and generalizability of algorithm

• Clinical reader performance
  – Assessment of clinical aids
  – Clinicians’ performance utilizing device
    • Multi-reader multi-case designs
    • Compare clinician’s performance with the ML SaMD aid to without the aid

5/9/2019
PROPOSED REGULATORY FRAMEWORK FOR AI/ML ALGORITHMS MODIFICATIONS

• Agency proposing framework to give manufacturers option to submit a plan for AI/ML modifications during initial premarket review

• Initial premarket phase would include
  – Review initial SaMD performance
  – Review plan for modifications
  – Review ability to manage/control resultant risks of modifications

• FDA asking for community feedback on this document

COMPONENTS OF CHANGE CONTROL PLAN

• Good ML Practices (GMLP):
  – Accepted practices in AI/ML algorithm design, development, training, and testing that facilitate the quality development and assessment of AI/ML-based algorithms
    • Based on concepts from quality systems, software reliability, machine learning, and data analysis, etc.

• SaMD Pre-Specifications (SPS):
  – Delineates the proposed types of modifications to the SaMD (i.e., what types of changes the sponsor plans to achieve)
    • Determine “range of potential changes” around the initial specifications and labeling of original device

• Algorithm Change Protocol (ACP):
  – Describes the methods for performing and validating the changes pre-specified in SPS (i.e. how the sponsor intends to achieve the changes)
    • Typically specific to the device and type of change
    • Expected to contain a step-by-step delineation of the procedures to be followed
CURRENT AI/ML WORKFLOW

Data selection and management

Model training and tuning

Model validation
  - Performance evaluation
  - Clinical evaluation

Log and track

Evaluate performance

Good Machine Learning Practices

Data for re-training

New (Live) Data

Deployed Model

Model monitoring
  - Log and track
  - Evaluate performance

Premarket Assurance of Safety and Effectiveness for Modified AI/ML algorithm

Legend

AI Model Development

AI Production Model

AI Device Modifications
PROPOSED TPLC APPROACH
OVERLAYED ON AI/ML WORKFLOW

Good Machine Learning Practices

Data selection and management

Model training and tuning

Model validation
  - Performance evaluation
  - Clinical evaluation

Data for re-training

Culture of Quality and Organizational Excellence

Premarket Assurance of Safety and Effectiveness

Review of SaMD Pre-Specifications and Algorithm Change Protocol

New (Live) Data

Deployed Model

Model monitoring
  - Log and track
  - Evaluate performance

Real-World Performance Monitoring

Legend

AI Model Development
AI Production Model
AI Device Modifications
Proposed TPLC Approach
1. How can current AI technology, particularly convolutional neural networks (CNN), be enhanced? Are there any gaps or roadblocks?

2. What additional tools beyond AI do we need …?
   – Develop data models & robust study designs for assessing standalone & clinical performance of AI/ML algorithms

3. How should the curated data sets be developed to optimize specific tasks of today and tomorrow?
   – Concurrently maximize impact and durability of datasets

4. What are the models for scaling up and sustainment? …
   – Develop data, study design and statistical methods for assessing AI/ML algorithm modifications
   – Determine elements for software pre-specifications (SPS) & algorithm change protocol (ACP)
   – Determine role of real-world role evidence in:
     • Assessing benefit of AI/ML algorithms
     • Supporting transparency to end user
<table>
<thead>
<tr>
<th>PI Name</th>
<th>Project Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>Technical and statistical assessment of AI/ML in digital pathology for clinical deployment</td>
</tr>
<tr>
<td>Gallas</td>
<td>High-throughput truthing of microscope slides to validate AI algorithms analyzing digital scans of same slides</td>
</tr>
<tr>
<td>Gavrielides</td>
<td>Improving pathologist performance for diagnosing ovarian cancer histological subtypes using ML</td>
</tr>
<tr>
<td>Glick</td>
<td>Development of deep learning model observer to assess x-ray breast imaging systems</td>
</tr>
<tr>
<td>Li and Petrick</td>
<td>Vascular calcium and material characterization in women using dual-energy CT (quantitative imaging biomarkers)</td>
</tr>
<tr>
<td>Pezeshk</td>
<td>Recurrent conv. networks for nodule detection in thoracic CT scans</td>
</tr>
<tr>
<td>Pezeshk, Graff</td>
<td>Comparison of quality assessment methods for deep-learning-based MR image reconstruction</td>
</tr>
<tr>
<td>Pezeshk</td>
<td>Assessment of AI systems that use un-annotated or weakly labeled datasets in training</td>
</tr>
<tr>
<td>Sahiner, Cha</td>
<td>Use of synthetic data in deep learning algorithm training and test</td>
</tr>
<tr>
<td>Sahiner, Chen</td>
<td>Leveraging imperfect post-market reference indices for the evaluation of artificial intelligence and machine learning devices</td>
</tr>
<tr>
<td>Sahiner, Gossmann</td>
<td>Assessment of adaptive machine learning systems: Methods for re-use of holdout sets and application to deep learning systems for medical image and physiological signal analysis</td>
</tr>
<tr>
<td>Zeng, Graff</td>
<td>Deep learning-based image reconstruction and denoising in radiological imaging</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

• I’d like to acknowledge Berkman Sahiner and Matthew Diamond for their help in developing this presentation