NevCon vs Material

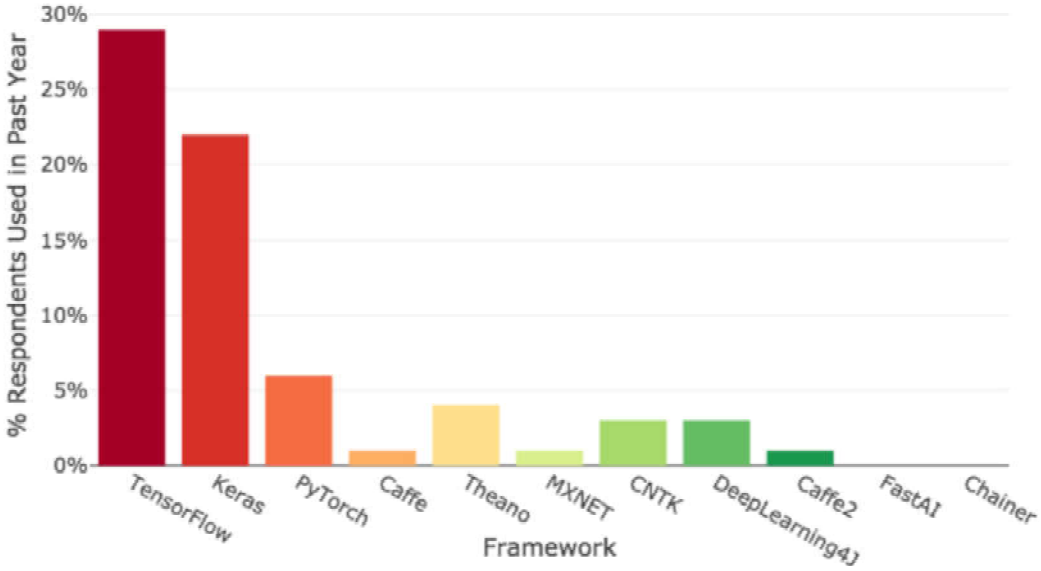
Sunday, 12 July 2020 4:42 pm

NevCon

* Nevigation
Localization – estimate the current user location
Path Planner – estimate the path from current location to destination/ keep going
Obstacle Avoidance – determine the walkable path based on Path planner

VSLAM (maps)
A* algorithm (Shortest distance)
GPS Module (Google Maps)
Obstacles
Adaptive Ground Segmentation
Optimal Walkable Direction Search

* Recognition
CNN Based Object Detection – detect the object category
Depth Based Object Detection – detect the object location and orientation



Keras
•Functional API
•Works like playing Lego bricks
•Multi-input, multi-output, arbitrary static graph topologies
•Good for 95% of use cases

Multiple inputs?
An video and a question

- "What is the person doing?"
  "Tidying"
- "Who is the person?"
  "Kondo"

Deep learning
Keras + Tensorflow
(Keras + Plaidml Currently deep learning mostly runs on Nvidia GPU)

Spatial vs sequential

	CNN	RNN / LSTM
Usage	Suitable for spatial data, e.g. image, video	Suitable for temporal data (sequential data), e.g. text, speech
Capability	Considered more powerful than RNN	Less powerful and slower in calculation
Input	Take fixed size inputs and generate fixed size outputs	Can handle arbitrary input/output lengths
Nature	Use local connectivity pattern (through 2D convolution)	Use time series information

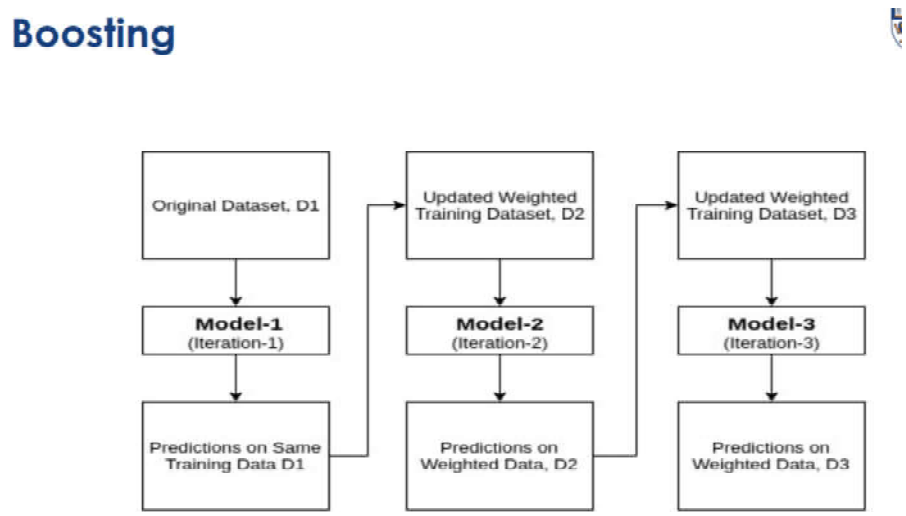
Create model diversity

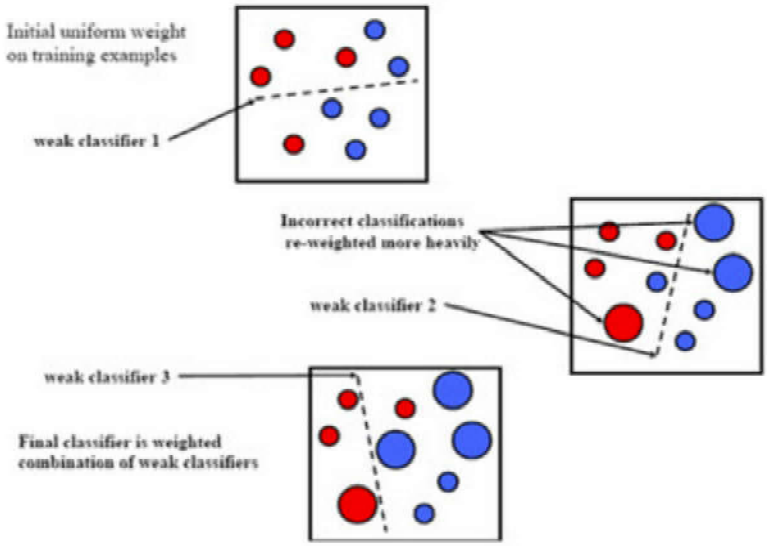
- Manipulating the training data (e.g. bagging)
- Manipulating the input features
- Varying the classifier type, architecture

Random Forests

Kinect's Decision Forest (3D) --> motion detection

Manipulate the dataset

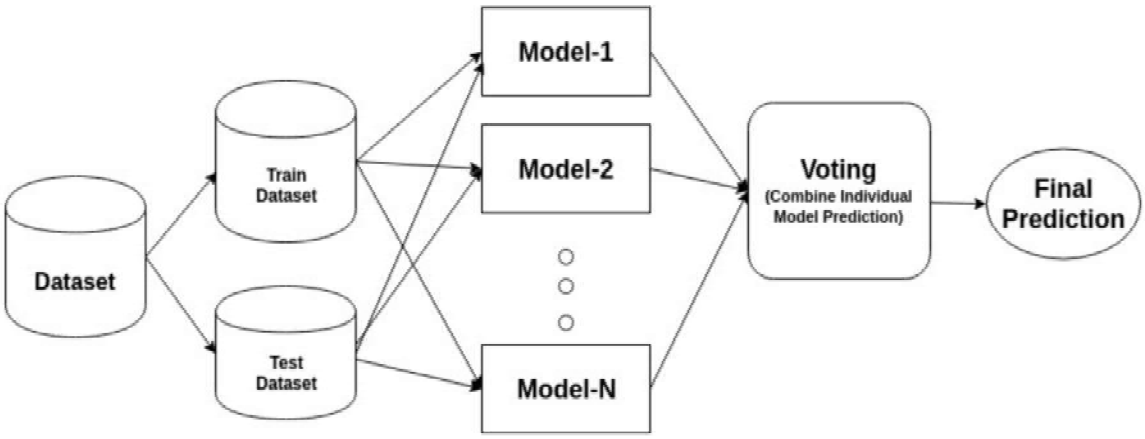




$$H(x) = sign(\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x))$$

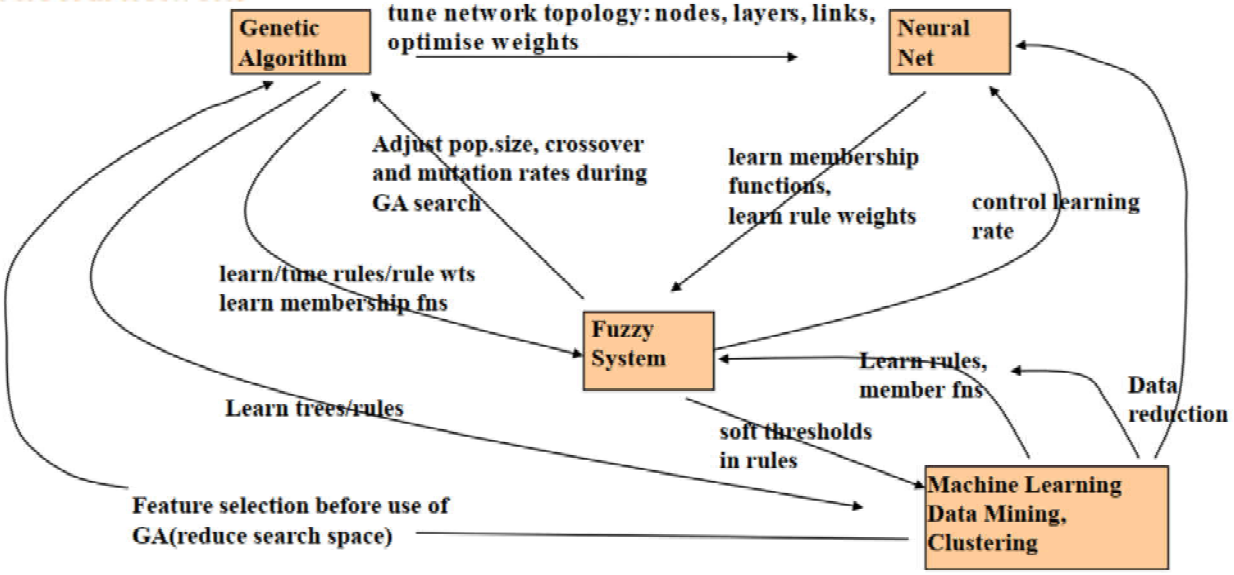
Hybrid system
Ensembel: weighted/unweighted

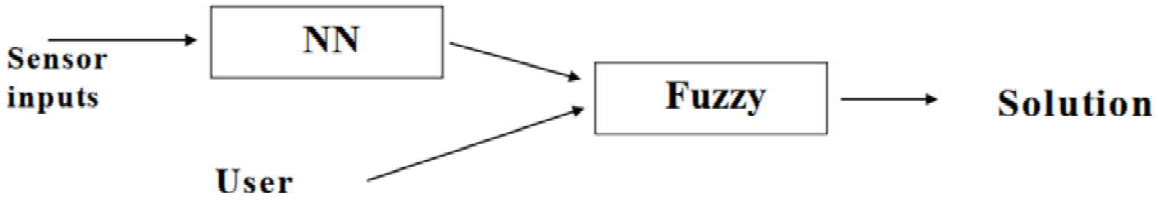
Ensemble

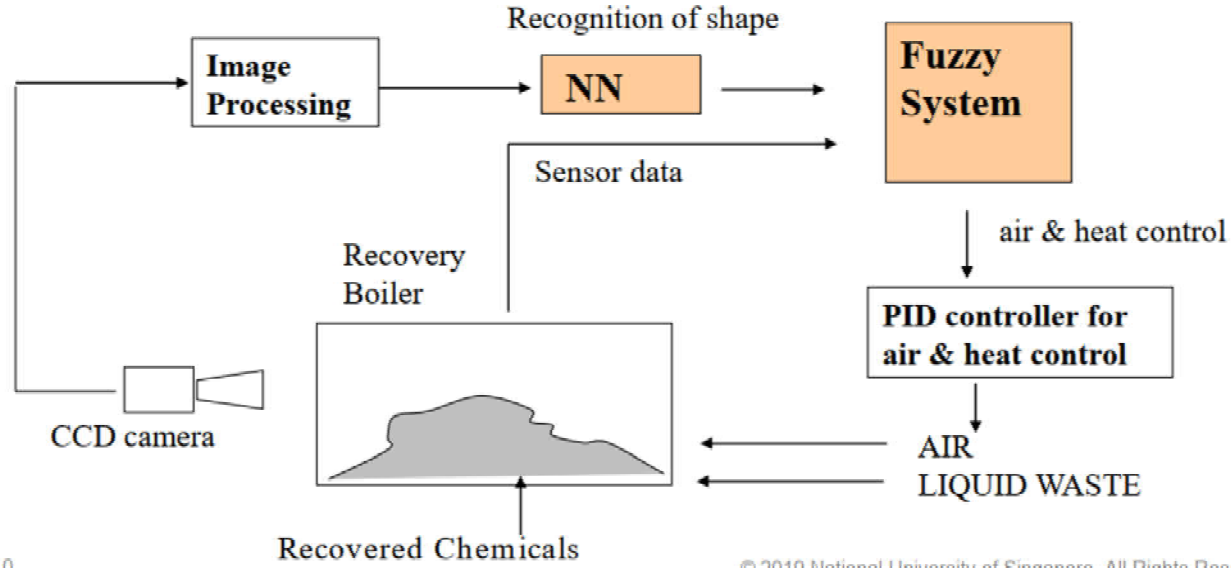


Self-tuning

- One technique is used to tune or learn the architecture for another,
  e.g: Neural network is used to learn a Fuzzy System , Genetic algorithm is used to optimise a neural network







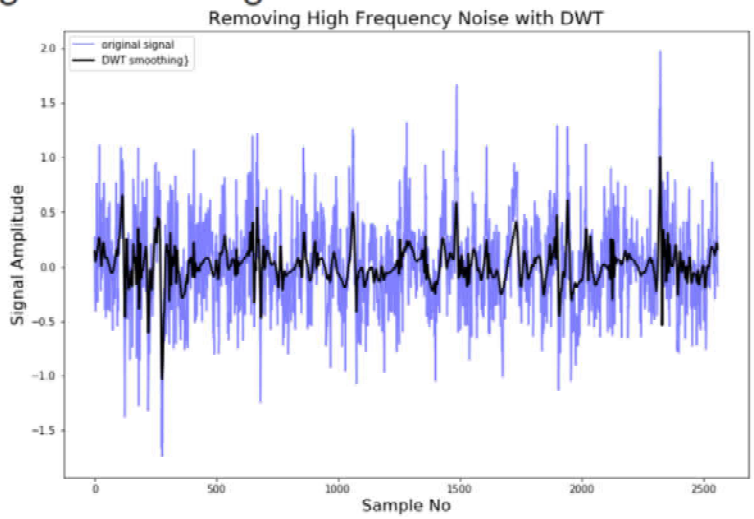
- Neural Network is used to represent and "learn" a Fuzzy System
- Nodes represent rule inputs, conditions, actions etc
- Special training algorithm required

*Human Machine Interaction
Speech and Audio
Text to speech
Speech Recognizer

Intelligent sensing
Naive bayse classifer
DWT: discrete wavelet transform (DWT) extracts meaningful information in a time-frequency domain

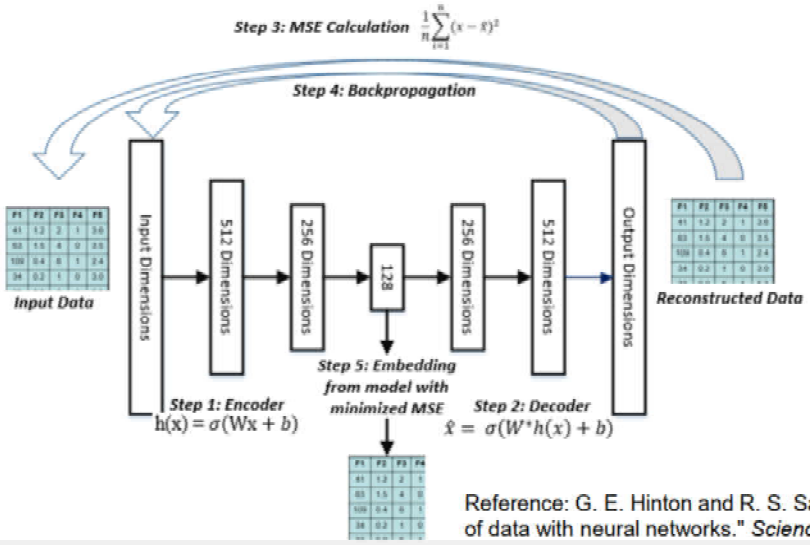
the common DWT-based feature selection follows a passive mode: manipulation through human experience or exhaustive trials

- Signal denoising



Signal representation learning can be deeper

- **Step 1:** Encoder "encodes" input data into a embedding using non-linear activation functions.
- **Step 2:** Decoder reconstructs output by using non-linear layers to "decode" embedding.
- **Step 3:** *Mean Squared Error* (MSE) is calculated between the reconstructed output and original input. Error is back-propagated to adjust autoencoder weights.
- **Step 4:** Steps 1-3 are repeated until MSE is minimized.



Reference: G. E. Hinton and R. S. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507

Connect vision and speech

Making sense of sensor data pipeline (5)

Contextual Relationship

- Understand the recognized data in context. The context can be obtained in multiple ways. These include using other sensors as well as using the history of recognized data for context recognition.
- For vision: A face recognition surveillance system can identify a potential invasion or normal operation depending on when the recognition takes place (day vs. night) or depending on the person that is recognized.

Semantic Relationship

- The connection between various recognized entities, for example the semantic relationship between a key and a lock refers to the operation of opening or closing the lock with the key
- For audio, the sentence "*I like coffee*" connects "I" (the subject) with "coffee" (object) using the predicate "like."
- For vision: semantic context of the video can be analyzed based on identified entities (multiple objects, locations, people, and activities) and scoring these entities with respect to their co-occurrence as well as relation to the type of classification scenario of importance.

Single type of sensor data: Audio

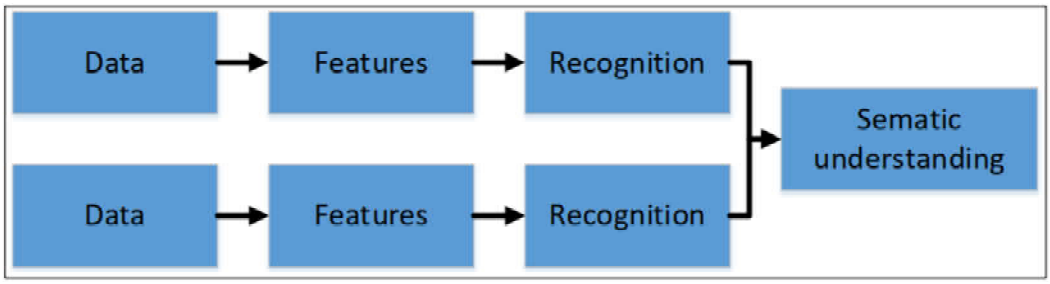
- Audio classification: A common IoT use case for microphones is to classify the environment that the sound was captured in.
- Voice activity detection: Another common use case is voice activity detection. Here the focus is on attempting to determine whether there is voice in the captured audio.
- Speaker recognition: Speaker recognition, sometimes also referred to as voice recognition, attempts to determine who is speaking.
- Keyword recognition: Recognize whether a particular word was uttered. Keyword recognition can also be generalized to keyphrase recognition and both are typically used as triggers for additional activity such as starting a session of commands or bringing up an application.
- Command and control: Command and control refers to using a small set of phrases in speech recognition. For illustration, this could include a set of commands to control a toy car such as "move forward," "move backward," "go faster," "go slower," "turn right/left," etc.

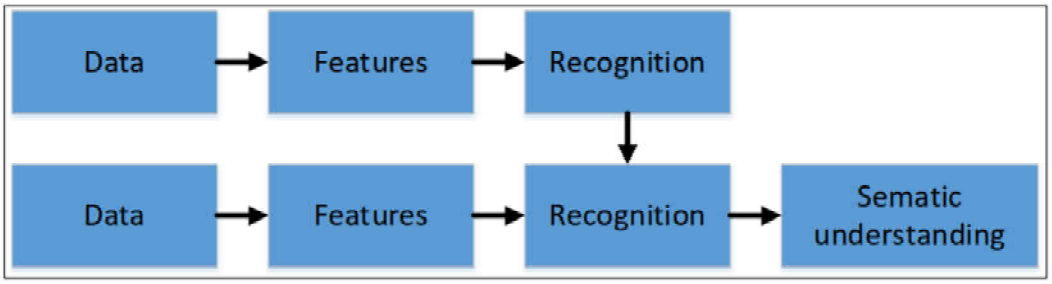
Urban sound 1302

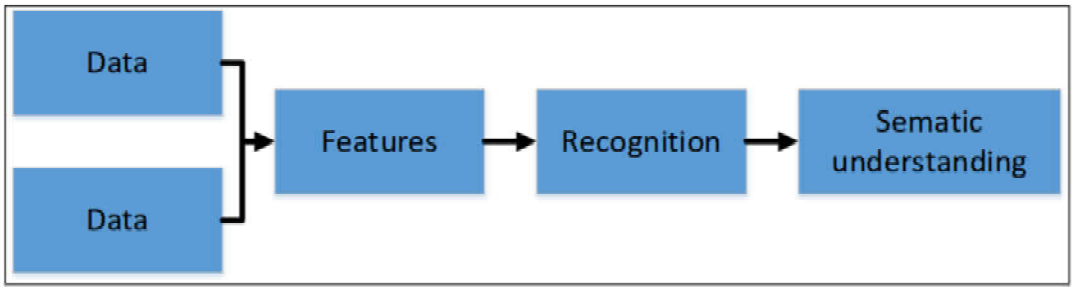
https://urbansounddataset.weebly.com/urbansound.html

Single type of sensor data: Vision

- Object recognition: Identify objects in an image and potentially matching them to a pre-existing database of objects that have been captured before.
- Face recognition: Detect a face in an image as well as matching that face against a database to label the face accordingly.
- Gesture recognition: Recognize static poses or moving gestures either specific to the hand/arm or the human body.
- Scene recognition: Identify multiple objects, faces, and people in an image and using that information to determine the likely activity or context.
- Anomaly detection: Identify if any anomaly occurred which should trigger additional analysis.
- Video summarization: Summarize the salient aspects of a long video stream. This includes scene changes, key scenarios and objects/characters that are the focus of the video.







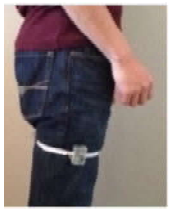
Re-orient

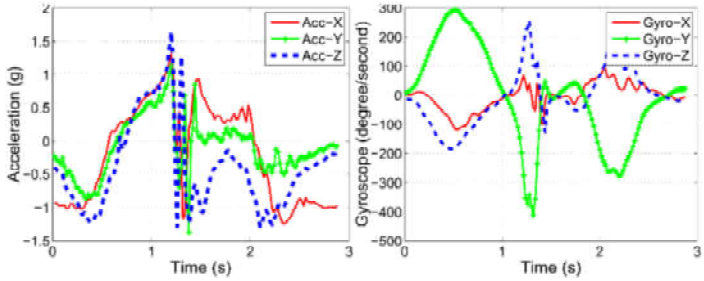
Single type of sensor data: Inertial

- An accelerometer essentially measures the force (proper acceleration) along x, y and z-axes.

Use cases

- Understanding of position/orientation helps mobile phones re-orient the screen in portrait or landscape mode and reverse direction as required.
- Gesture recognition based on buffering continuous data and looking at the change in force and orientation.





Project Examples:

- Stock market forecasting
- House price prediction
- Diabetes occurrence prediction
- Object recognition
- Face recognition
- Automatic image caption generation
- Wearable sensors for human activity recognition
- Hand gesture recognition
- ……

- The project must develop, integrate and demonstrate at least **three** out of following aspects:
  - ❖ Supervised learning / unsupervised learning scenarios
  - ❖ Machine learning/ Deep learning techniques
  - ❖ Hybrid machine learning /Ensemble approach
  - ❖ Intelligent sensing / sense making techniques