Generalized Surrogate Loss Function for Plant Health Detection

Dr. Hilman Ferdinandus Pardede

Research Center for Informatics, Indonesian Institute of Sciences

Presented at ASEAN IVO Forum 2019

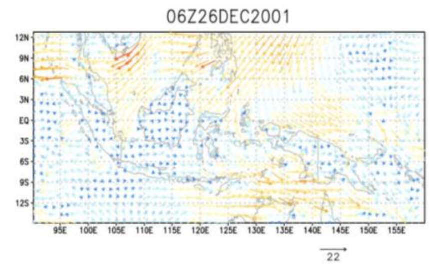
Manila, November 20, 2019

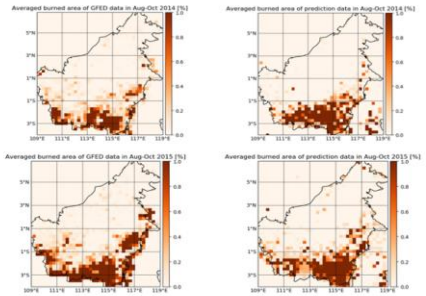
Research Activities at Research Center for Informatics - LIPI

Research Areas:

- High performance computing
- Computer visions
- Data Security
- Information retrieval
- Machine Learning and Pattern Recognition

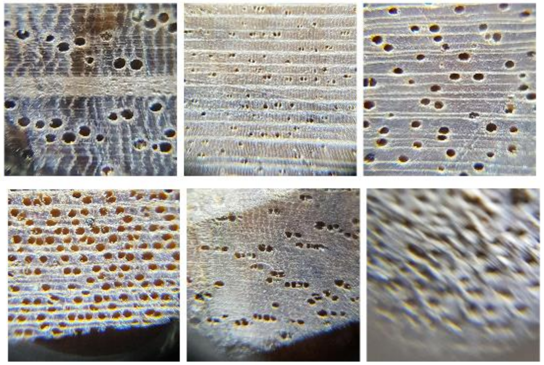
Parallel computing and climate modeling

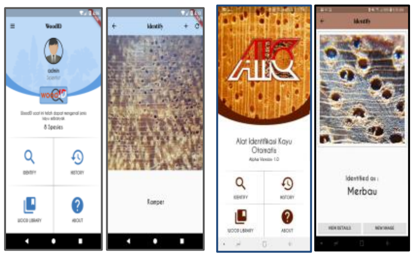




Contact: **Dr. Zaenal Akbar (zaenal.akbar@lipi.go.id)**

Wood identifications and smart parking





Contact: **Dr. Esa Prakasa (esap001@lipi.go.id)**

Robust watermarking for copyright protection

Watermarked

Extracted Host

Extracted Watermark

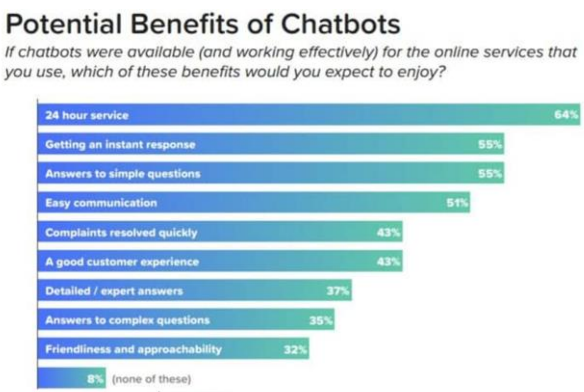
Contact: **Dr. Didi Rosiyadi (didi016@lipi.go.id)**

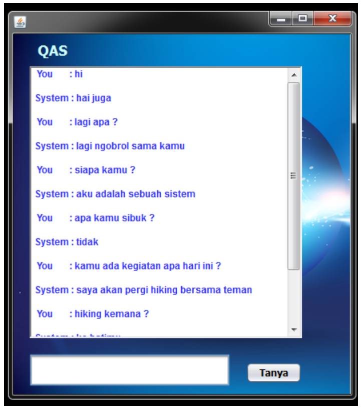
Research Activities at Research Center for Informatics - LIPI

Research Areas:

- High performance computing
- Computer visions
- Data Security
- Information retrieval
- Machine Learning and Pattern Recognition

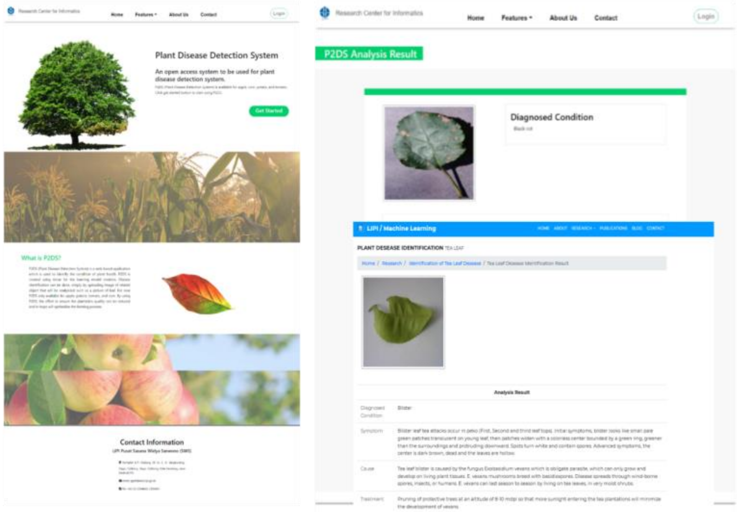
Q-A systems for healthcare

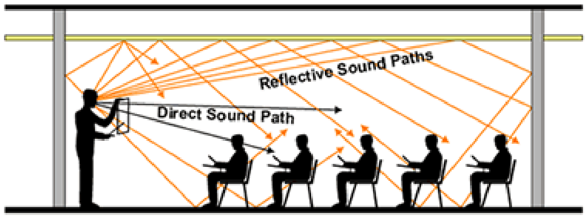




Contact: **Dr. Wiwin Suwarningsih (wiwi007@lipi.go.id)**

Object recognition and speech recognition

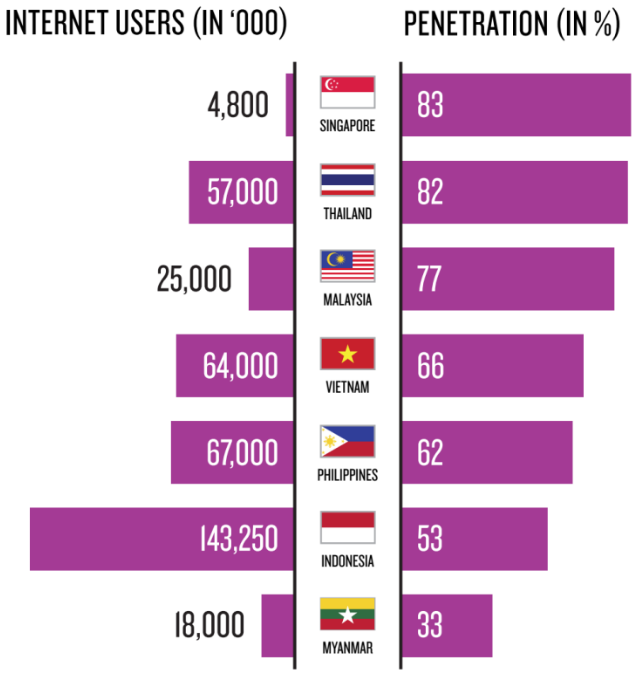




Contact: **Dr. Hilman Pardede (hilm001@lipi.go.id)**

Background

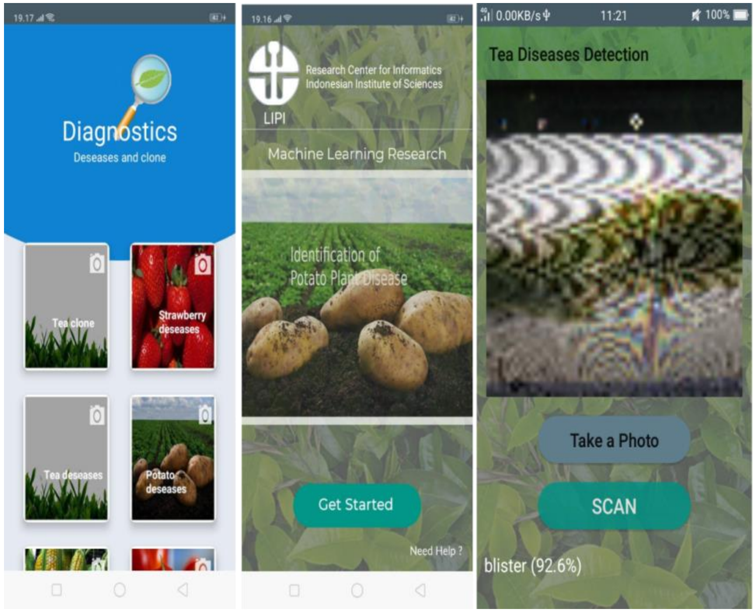
1. Machine learning and data mining technologies have penetrated many areas from finance, legal, agriculture, health, security, etc
    - Smart Farming and Precision Agricultures
2. Involvement of ICT technologies on Farming in ASEAN Countries, in particular in Indonesia, is still very limited
    - Agriculture is a pillar of ASEAN economy
    - ASEAN ICT Masterplan 2015 (ASEAN, 2011), ICT will become an engine of growth for ASEAN
3. Plant diseases detection for tea (Grant from Indonesian Ministry of Research and Technology)
    - **20 to 40 %** of global crop losses because of plant diseases (Source FAO)
    - Tea is one major agricultural products in some ASEAN countries



Internet Penetration (Nielsen, 2019)

Background

No	Types of Diseases	#data
1	Healthy	1,448
2	Blister blight	1,104
3	Empoasca	3,727
4	Caterpillars attack	1,616
5	Yellow-mite	2,484
6	Helopeltis	2,240
Total		12.619



Healthy Blister Blight Empoasca Helopeltis Caterpillar YellowMite

Condition	AlexNet	GoogleNet
Original	81.62	77.30
Blurred	63.78	65.05
Rotated(10^{o})	79.19	62.79

Current development:

- Supervised and unsupervised deep learning.
- Mobile and web frameworks
- Robustness issues.

Challenges:

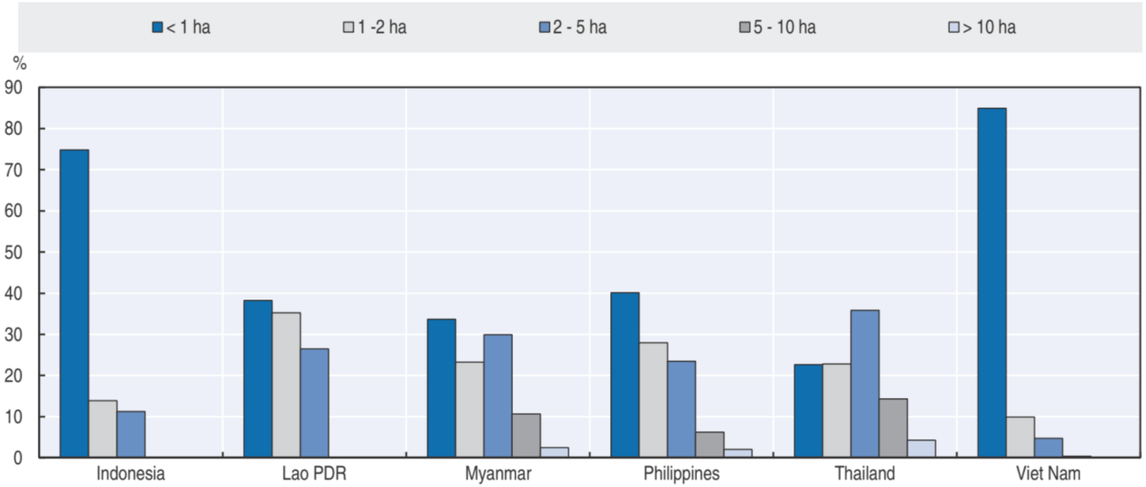
- Performance and robustness issues
- Lacks of infrastructures
- Lack of data/resources
- Limitations of Current state-of-the-art techniques: Deep Learning

Problem Statements

1. Deep Learning (DL) is currently state-of-the-art of AI technologies for complex data such as image, speech, and video.
2. Surrogate Loss Function (SLF) to replace non-tractable 0-1 loss function that is used in conventional ANN motivates wide DL implementations
3. DL systems are still not robusts
4. Current SLFs do not quite approximate the 0-1 loss function especially positive values which tend to infinity
5. Tsallis entropy:
   - Generalization of Boltzmann-Gibbs and Shannon entropy to deal with *complex* data/systems.
   - Under this framework, better statistical solutions could be derived.
6. **Tsallis statistical frameworks could be utilized to generalized SLFs**
   - Closer or more relaxed than 0-1 loss

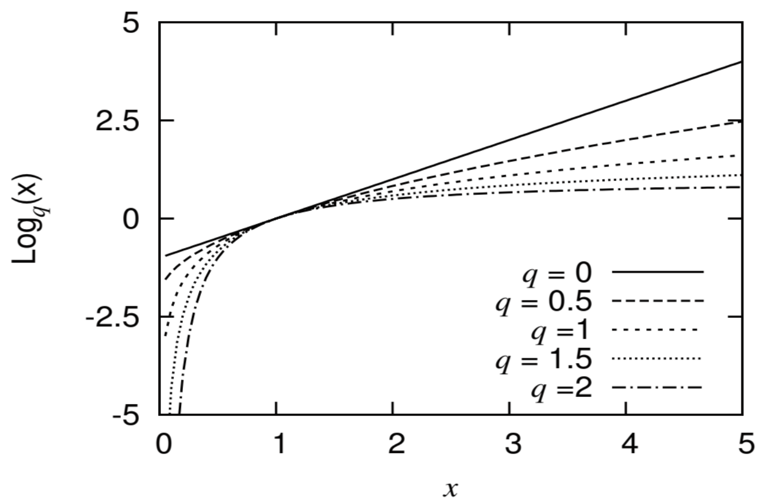
Targets

1. Develop better and more robust tunable surrogate loss function for classification
   - Evaluation the optimization formula from the loss function for classification tasks: plant health detection systems
2. Develop plant health detection systems based on the proposed loss function
   - Available for smallholder farmers → difficult access to automatic tea diseases detection
3. Collaborations for 2020 ASEAN IVO Call for Proposals
   - Knowledge capacity
   - Resource sharing



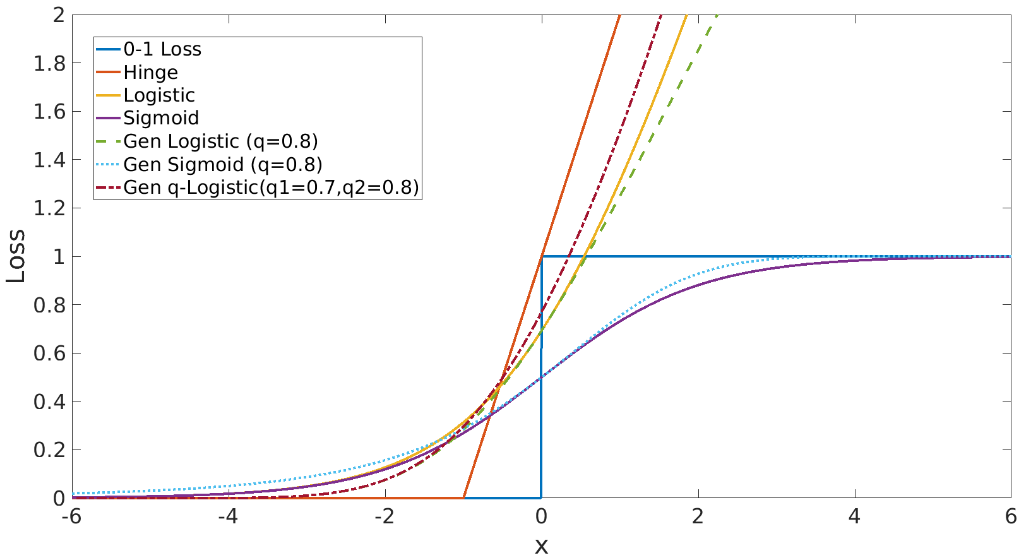
Distribution of farm size in ASEAN Countries (FAO, 2017)

Proposed Method



The q-log functions: $$\log_q(x) = \frac{x^{1-q} - 1}{1-q}$$

The q-exp functions: $$\exp_q(x) = (1 + (1-q)x)^{\frac{1}{1-x}}$$



$$S(z) = \log(1 + \exp(z))$$ Logistic

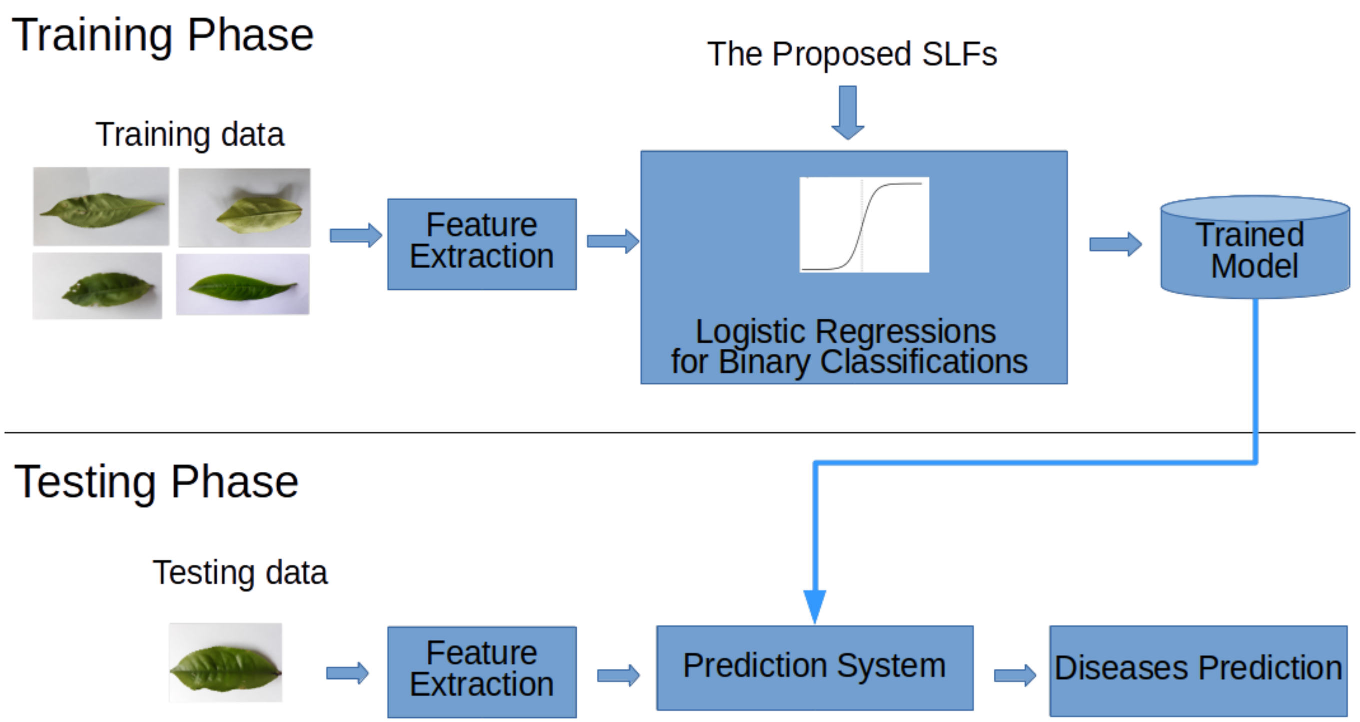
$$S_q(z) = \log(1 + \exp_q(z))$$ Gen Logistic

$$S_q(z) = \log_q\ (1 + \exp_q(z))$$ Gen q-Logistic

$$S(z) = \frac{1}{1 + \exp(-z)}$$ Sigmoid

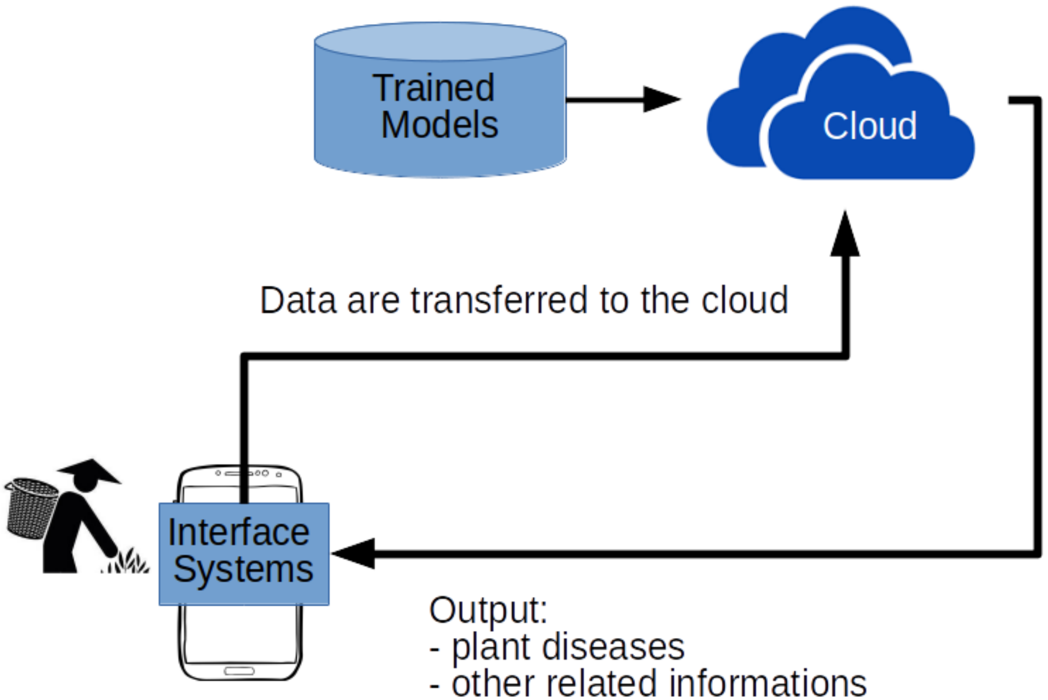
$$S_q(z) = \frac{1}{1 + \exp_q(-z)}$$ Gen Sigmoid

Proposed Method



1. Evaluate the proposed SLFs on simple logistic regression-based classifications
2. Test the systems on plant health detection data
3. Extend the systems on Deep Learning systems

Proposed Method



4. Develop applications for plant health detection that are accurate and robust
   - Provide tools for small-holder farmers for plant conditions
   - Drive a cultural change of implementing ICT technologies for farming in ASEAN Countries

Collaborations

- Prospective Partners in ASEAN countries and Japan
- Capacity in machine learning techniques
  - The Center for Information and Neural Networks (CiNet), NICT
    - Algorithms development
    - Capacity development in Deep learning technologies
- Data sharing
  - Universiti Putra Malaysia: Hyperspectral imaging for plant diseases detection
  - University of Malaya: Deep Learning for Plant Species Classification
  - Ho Chi Minh City University of Technology: deep learning techniques and UAVs for paddys field assessment
  - Hosei University: Viral Plant Diseases Using Convolutional Neural Networks
  - RIKEN Center for Sustainable Resource Science: Machine learning for phenotyping of plant for improving its productivity
  - Plants or data local to ASEAN Countries: Oil palm, rubber, coffee
- Resource Sharing
  - Computing facilities
  - Software development



Outputs and Outcomes

1. New surrogate loss functions derived from Tsallis Statistics
   - More accurate and robust classifiers in particular for plant health detection systems
2. A new dataset of agriculture commodities that are unique for ASEAN countries
   - More variants of diseases that are local to each ASEAN Countries.
   - Encouraging similar projects in the area of machine learning and data mining for agriculture
3. A robust web-based and/or Android based applications for Plant Health/Plant diseases detection
   - Provide ICT-based technologies / tools for small-holder farmers
   - Support the implementation of smart-farming and precision agriculture
4. Collaborations
   - Improving researchers capacity
   - Resource and data sharing
     - Tea is major commodity in Malaysia and Thailand

Conclusion

1. Farming in Indonesia still heavily rely on traditional approaches
    - Involvement of ICT technologies is still very limited
2. Plant health/plant diseases detection is one needed applications for agricultural products
    - One major cause of crop failures
3. Deep learning is state-of-the-art for plant health detections
    - Robustness Issues
4. Robust and more accurate classifiers using Surrogate Loss Functions derived under Tsallis Statistics
5. The availability of robust and accurate plant health detections may help smallholder farmers
6. The dataset availability may drive further research in the area
7. Look for partners that are interested in the area
    - Smart farming/precision agriculture
    - Machine Learning and Deep Learning

Thank You

Contact:
Hilman F. Pardede
hilman.f.p@ieee.org, hilm001@lipi.go.id, hilman.fp@gmail.com