iiWAS 2019 MoMM 2019

The 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS2019)

The 17th International Conference on Advances in Mobil Computing & Multimedia (MoMM2019)

December 2-4, 2019
Munich, Germany

PROGRAM**BOOK**

CONTENT

Welcome Message

Gabriele Anderst-Kotsis
iiWAS/MoMM2019
General Conference Chair

Ismail Khalil
President @WAS

Maria Indrawan-Santiago
iiWAS2019 Program
Committee Co-Chair

Eric Pardede
iiWAS2019 Program
Committee Co-Chair

Pari Delir Haghighi
MoMM2019 Program
Committee Chair

Welcome to the 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS2019) and the 17th International Conference on Advances in Mobile Computing & Multimedia (MoMM2019).

Until today Munich, Germany was best known for its world-renowned Oktoberfest celebration of fine food, music and fermented hops.
But from today forward, Munich welcomes you to iiWAS/MoMM 2019, the premier international forum for like-minded academics and innovators, meeting to share digital discoveries, new methodologies, and ground-breaking applications of our core technologies. In a manner of speaking, it's our own personal **DecemberFest**.

As iiWAS/MoMM has done for over 20 years, this exciting exchange of ideas will be delivered by an impressive array of international participants and additional distinguished speakers focused on providing you three distinct levels of experiences:
A Learning Experience featuring a very strong program of over 100 thought-provoking presentations and keynote speeches in over 25 sessions.
A Networking Experience, offering you valuable opportunities to form lasting friendships and productive collaborations with like-minded academic leaders from around the globe, as you'll find among this year's participants coming from 36 countries and over 150 prestigious research and academic institutions.
A Motivational Experience, exposing the young, upcoming colleagues among us to exciting new innovations in our fields as a springboard to motivate their own research and discoveries.
We also have a very enjoyable Social program this year, which features a welcome reception at LMU Lichthof and a gala dinner at Augustiner Keller München where you will enjoy a traditional Bavarian Dinner. The Lagerkeller is reserved for iiWAS/MoMM 2019 Conference Dinner exclusively and you will have the opportunity to taste Bavarian beer as well as hearty local food.

As the conference organizers, we would like to express our sincere thanks to all members of the program committees for carefully reviewing all paper submissions and for putting together this interesting program. We are convinced that you will be inspired by the presentations and talks constituting this year´s programme and hope that this will also lead to subsequent cooperation.

We are very grateful to our distinguished keynote speakers: Assoc. Prof. Flora Salim from CIDDA, RMIT University, Melbourne, Australia; Prof. Albrecht Schmidt from LMU Munich, Germany and Prof. Dirk Draheim from Tallinn University of Technology, Estonia. We are also very grateful for our panel chair Prof. Alfredo Cuzzocrea from the University of Calabria, Italy.

We tried our best in arranging everything locally for your comfort. These people and many others behind the scenes created this special spirit which makes all of us feel like coming home every year when we meet at iiWAS and MoMM.

So, welcome to Munich for iiWAS/MoMM 2019. And thank you for your participation, we're confident you will find the conference both fulfilling and rewarding.

Committees

General Conference Chair
Gabriele Anderst-Kotsis, Johannes Kepler University Linz, Austria

Program Committee Chairs
(iiWAS2019)
Maria Indrawan-Santiago, Monash University, Australia
Eric Pardede, La Trobe University, Australia
(MoMM2019)
Pari Delir Haghighi, Monash University, Australia

Steering Committee Co-chairs
Ismail Khalil, Johannes Kepler University Linz, Austria
David Taniar, Monash University, Australia
Syopiansyah Jaya Putra, UIN Syarif Hidayatullah Jakarta, Indonesia

Proceedings Co-Chairs
Matthias Steinbauer, Johannes Kepler University Linz, Austria
Ivan Salvadori, Federal University of Santa Catarun, Brazil

Communcations Chair
Paul Wiesinger, Johannes Kepler University Linz, Austria

General Secretary
Hesti Sudjana, @WAS Organization

Program Committee iiWAS2019

Ali Alwan, International Islamic University Malaysia, Malaysia
A. S. M. Kayes, La Trobe University Australia
Abdelkader Hameurlain, IRIT - Paul Sabatier University, France
Akiyo Nadamoto, Konan University Japan
Brahim Ouhbi, ENSAM, Morocco
Christine Bauer, Johannes Kepler University Linz, Austria
Christine Strauss, University of Vienna, Austria
Dan Johansson, Umea University, Sweden
Devis Bianchini, Universita' degli Studi di Brescia, Italy
Dimitris Karagiannis, University of Vienna, Austria
Dion Goh, Nanyang Technological University, Singapore
Edgar Weippl, SBA Research, Austria
Erich Schikuta, University of Vienna, Austria
Ester Zumpano, University of Calabria, Italy
Hamidah Ibrahim, Universiti Putra Malaysia, Malaysia

Committees

Harald Wahl, University of Applied Sciences Technikum Wien, Austria
Haruo Yokota, Tokyo Institute of Technology, Japan
Hind Lamharhar, Mohamed V Agdal University, Mohammadia School of Engineers(EMI), FMD, Morocco
Hui Ma, Victoria University of Wellington, New Zealand
Jeff Tan, IBM Research, Australia
João Moura-Pires, DI FCT – UNL, Portugal
Johann Eder, Alpen Adria Universität Klagenfurt, Austria
Johannes Sametinger, Johannes Kepler University Linz, Austria
Kenji Hatano, Doshisha University, Japan
Ladjel Bellatreche, LIAS/ENSMA, France
Lorna Uden, Staffordshire University, UK
Luca Mazzol, Hochschule Luzern (HSLU), Switzerland
Manolis Gergatsoulis, Ionian University, Greece
Marco Mesiti, DICO - University of Milano, Italy
Marco Antonio Casanova, PUC – Rio, Brazil
Martin Gaedke, Chemnitz University of Technology, Germany
Masayoshi Aritsugi, Kumamoto University, Japan
Mayumi Ueda, Kyoto University, Japan
Michele Melchiori, Università degli Studi di Brescia, Italy
Muhammad Younas, Oxford Brookes University, UK
Naoko Kosugi, Tokyo Healthcare University, Japan
Norwati Mustapha, University Putra Malaysia, Malaysia
Raymondus Kosala, BINUS International - BINUS University, Indonesia
Rodica Potolea, Technical University of Cluj-Napoca, Romania
Rodolfo F. Resende, Federal University of Minas Gerais, Brazil
Ronaldo Mello, Federal University of Santa Catarina, Brazil
Sami Habib, Kuwait University, Kuwait
Sherif Sakr, The University of New South Wales, Australia
Sucha Smanchat, King Mongkut's University of Technology North Bangkok, Thailand
Takahiro Hara, Osaka University, Japan
Takako Hashimoto, Chiba University of Commerce, Japan
Toshiyuki Amagasa, University of Tsukuba, Japan
Valeria De Antonellis, Università degli Studi di Brescia, Italy
Wanita Sherchan, IBM, Australia
Werner Winiwarter, University of Vienna, Austria
Wookey Lee, Inha University, South Korea
Yacine Sam, University François Rabelais – Tours, France
Yousuke Watanabe, Nagoya University, Japan
Zaher Al Aghbari, University of Sharjah, UAE

Committees

Program Committee MoMM2019

Ajith Abraham, MIR Labs, USA
Chow-Sing Lin, National University of Tainan, Taiwan
Tzung-Shi Chen, National University of Tainan, Taiwan
Stamatis Karnouskos, SAP
Agustinus Borgy Waluyo, Monash University, Australia
Andreas Schrader, University of Luebeck, Germany
Andrzej Romanowski, Technical University of Lodz, Poland
Antonio Liotta, Eindhoven University of Technology, Netherlands
Bin Guo, Institut Telecom SudParis, France
Carlos Calafate, Universitat Politècnica de València, Spain
Chang Wu Yu, Chung Hua University, Taiwan
Charles Gouin-Vallerand, Télé-Université du Québec, Canada
Chien-Cheng Lee, Yuan Ze University, Taiwan
Chun-Hsin Wu, National University of Kaohsiung, Taiwan
Clemens Holzmann, University of Applied Sciences Upper Austria, Austria
Fernando Zacarias Flores, Universidad Autonoma de Puebla, Mexico
Francesco Piccialli, University of Naples Federico II, Italy
Hermann Hellwagner, Klagenfurt University, Austria
Hong Va Leong, Hong Kong Polytechnic University, China
I-Hsien Ting, National University of Kaohsiung, Taiwan
Jaakko Hakulinen, Tampere University, Finland
Juan Ye, University of St Andrews, UK
Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan
Keith Cheverst, Lancaster University, UK
Klaus Turowski, Otto von Guericke University Magdeburg, Germany
Markku Turunen, University of Tampere, Finland
Natalia Kryvinska, University of Vienna, Austria
Panayotis Fouliras, University of Macedonia, Greece
Paolo Bellavista, University di Bologna, Italy
Roman Kolcun, Imperial College London, UK
Sami Habib, Kuwait University, Kuwait
Sara Comai, Politecnico di Milano, Italy
Sherali Zeadally, University of Kentucky, USA
Simon Mayer, TU Graz and Pro2Future GmbH, Austria
Svetlana Boudko, NR (Norwegian Computing Center), Norway
Takahiro Hara, Osaka University, Japan
Tommi Mikkonen, University of Helsinki, Finland
Tomoki Yoshihisa, Osaka University, Japan
Tsutomu Terada, Kobe University, Japan
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Vilmos Simon, Budapest University of Technology and Economics, Hungary
Wen-Yang Lin, National University of Kaohsiung, Taiwan
Whai-En Chen, National Ilan University, Taiwan
Wolfgang Schreiner, Johannes Kepler University Linz, Austria
You-Chiun Wang, National Sun Yat-Sen University, Taiwan
Young-Koo Lee, Kyung Hee University, South Korea
Yusuke Gotoh, Okayama University, Japan

General Information

Conference Venue
iiWAS2019, MoMM2019 and SHOW are endorsed by the International Organization for Information Integration and Web-based Applications & Services (@WAS) and supported by Institute for Telecooperation JKU Linz and Ludwig Maximilian University München, will be held at Holiday Inn Munich - City Center, that located not far from Munich Old Town.
Hotel Address:
Hochstrasse 3, Munich 81669, Germany
https://www.hi-hotel-muenchen.de/en/

Registration Desk
The Registration Desk is located at the Foyer Ballsaal (Ballroom)
All conference participants shall present themselves at the registration Desk for conference check in and collect the participant's packages.
Opening hours:
Sun, 1 Dec; 17.00 – 19.00 Mon, 2 Dec; 08.00 – 16.00 Tue, 3 Dec; 08.30 – 16.00

Coffee Break & Lunches
Coffee Break will be served at Forum 9-11, Ground floor
Morning Coffee Break between 10.30 – 11.00 and Afternoon Coffee Break at 15.30 – 16.00
Lunch will be served at the Restaurant , Ground floor, start at around 12.30 – 13.30.
Please always wear/bring your ID badge throughout the conference programs, coffee breaks and lunches.

Smoking Policy
Holiday Inn Munich - City Center is a smoking free place. Smoking only permitted in the restricted area and outside the building.

Mobile Phone Courtesy
iiWAS2019 & MoMM2019 requests that all mobile phones and other equipment with audible alarms be turned off in all sessions as a courtesy to the presenters and to the other attendees.

Presentation Equipments
The conference organizer will provide a standard presentation equipment i.e. Projector and flipchart. All presenters are recommended to bring your personal laptop for presentation including any adaptors required for your laptop, otherwise please inform the organizer earlier before your presentation schedule.

WLAN Access
For hotel guests, free WiFi is available in the entire hotel premises. For those who are not staying in Holiday Inn Munich - City Center, please contact registration desk for WiFi connection & availability in the main meeting room.

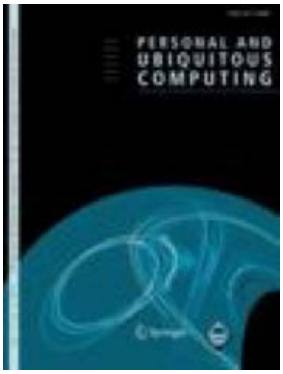
Electricity Supply
Electricity in Germany is supplied at 220V. Please make sure to bring your own travel adapter if needed. We cannot guarantee the availability of the appropriate adaptor or transformer.

Journal Special Issues

International Journal of Web Information Systems (IJWIS)

Journal of Data Intelligence (JDI)

International Journal of Pervasive Computing and Communications (IJPCC)



Journal Personal and Ubiquitous Computing (JPUC)

Keynote Speakers

Flora Salim
CIDDA, RMIT University, Melbourne, Australia

Context and Behaviour Modelling of User and Urban Activities

Abstract
Context is the most influential signal in analysing human behaviours. Effective and efficient techniques for analysing contexts inherent in the spatio-temporal sensor data from the urban environment are paramount. It is important to observe and learn the context from which the data is generated in, particularly when dealing with heterogenous high-dimensional data from individuals and groups of users, generated either from personal devices or sensors embedded in buildings, cities, and urban areas.
One main challenge in modelling user activities in urban areas is to discover meaningful correlations among the numerous sensor channels and other types of data from multiple domains. Data-driven models need to also reflect on the personalised behaviours of different users. On the other side of the coin, often data is generated by many companies without any information of individuals. However the analysis still needs to reflect different activity profiles.

I will present our generic temporal segmentation techniques that we have used for multiple applications. A new concept of cyber, physical, social contexts will be introduced, and how they translate in various domain applications of our research for analysing occupant behaviours, and for personalised intelligent assistants.
Finally, the data-driven models can be used as input for recommender systems for individuals and/or groups, and for optimising the routes and/or itineraries in multiple applications, from visitors in a shopping mall, to travelling parking inspectors and police officers.

Biography
Flora Salim is an Associate Professor at the Computer Science and IT (CSIT) discipline, School of Science, RMIT University, Melbourne, Australia, and a Deputy Director of the RMIT Centre for Information Discovery and Data Analytics (CIDDA). She leads the Context Recognition and Urban Intelligence group in CIDDA. A/Prof. Flora's research interests include context-aware situational intelligence and spatio-temporal analysis, specifically in human mobility tracking, context and behaviour modelling, time-series prediction, and cross-domain data fusion and learning. She obtained her PhD from Monash University in 2009.
She has received numerous awards for her work, including the Humboldt-Bayer Research Fellowship by Alexander von Humboldt Foundation and the Bayer Science & Education Foundation in 2019, Humboldt Fellowship for experienced researcher, Victoria Fellow 2018 (awarded by the Victorian government), the RMIT Vice-Chancellor's Award for Research Excellence – Early Career Researcher 2016; the RMIT Award for Research Impact - Technology 2018; Victorian iAwards (2014), Australian Research Council (ARC) Postdoctoral Research Industry Fellow (2012-2015); IBM Smarter Planet Industry Skills Innovation Award (2010); and a Google Anita Borg Scholar (2008).
Prior to her faculty appointment in CSIT, she was a postdoc in the Spatial Information Architecture Lab, part of the RMIT School of Architecture and Design, working on multiple design modelling and optimisation problems in architecture, engineering, and construction projects. Prior to her PhD, she was a Senior Software Engineer in mediaproxy, developing realtime signal and content monitoring solutions for TV & broadcasting companies.
She is an Associate Editor of the PACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) and an Area Editor of Pervasive and Mobile Computing.

Keynote Speakers

Albrecht Schmidt
LMU Munich, Germany

Making Humans Smarter Than Artificial Intelligence: Digital Technologies to Amplify Human Perception and Cognition

Abstract

Is artificial intelligence (AI) taking over the world? Are humans losing out to AI in the work place?
We are at a point in history, where it seems feasible that we create technologies that could become smarter than humans. This raises the fundamental question of what roles humans play in a future world. I argue that by using these advances in technologies and in particular in sensing and AI we can amplify human perception and cognition to levels that we could never before in history even imagine.
We are at the beginning of this fundamental transformation: the use of digital tools to amplify the mind. Many products, ranging from mobile access to search engines, to wearable devices for lifelogging and augmented realty application give us first indications of this transition.
In our research, we create novel digital technologies that systematically explore how to enhance human cognition and perception. Our experimental approach is to: first, understand the users in their context as well as the potential for enhancement. Second, we create innovative interventions that provide functionality that amplifies human capabilities. And third, we empirically evaluate and quantify the enhancement that is gained by these developments.
It is exciting to see how ultimately these new ubiquitous computing technologies have the potential for overcoming fundamental limitations in human perception and cognition and lead the way for creating cognitive and perceptual super powers. If we succeed to create technologies that seamlessly amplify human cognition and perception, humans augmented with these technologies will outsmart artificial intelligence.

Biography

Albrecht Schmidt is a professor for User Centered Ubiquitous Media in the Computer Science Department at the Ludwig-Maximilians Universität München (LMU Munich). He studied computer science in Ulm, Germany and Manchester, UK and receive in 2003 a PhD from the Lancaster University in the UK. For the last 15 years Albrecht has been dedicated to creating usable interactive systems and he coined the term “implicit interaction”.
The focus of his current work is on novel user interfaces to enhance and amplify human cognition. He is working on interaction techniques and interactive applications in the context of mobile and ubiquitous computing.
He is co-founder of the ACM conference on Tangible and Embedded Interaction (TEI), initiated the ACM conference on Automotive User Interfaces and co-chaired the ACM SIGCHI program in 2014. He is on the editorial board of ACM ToCHI and edits a forum on interaction technologies in the ACM Interactions magazine. In 2016 Albrecht Schmidt received a ERC Consolidator Grant to work on the Project “AMPLIFY: Amplifying Human Perception Through Interactive Digital Technologies”.

Keynote Speakers

Dirk Draheim
Tallinn University of Technology, Estonia

Collective Intelligence Systems from an Organizational Perspective

Abstract
In this talk, we consider Collective Intelligence (CI) systems from an organizational perspective. CI systems offer a solution to problems that need cognitive skills, problem-solving capabilities, knowledge, know-how or experience at large scale. They help to facilitate and streamline large-scale problem-solving endeavours. The organizational perspective on CI systems offers us two strands of discussion. On the one hand, it can be about understanding the potential of CI systems for today's organizations. On the other hand, CI systems can be considered as organizations themselves and can be investigated as such.
We start by reviewing the state-of-the art of CI frameworks. What are the essential building blocks of a CI system? Who uses them? For what, how and why? We come up with a generalized framework that serves us as a basis for further investigations.
From a governance perspective, today's organizations are recursive-feedback control systems, usually expressed in the form of process-oriented management. A deeper look reveals a plethora of different styles of organizational culture. Still, viable organizations have in common certain essential sub systems, which are policy making, external and internal steering, the primary activities and an informational backbone. How can we exploit CI systems to support these organizational building blocks? Can CI systems be made an integral part of organizations to make them more stable towards distortions; more adaptive towards an ever-changing environment; more agile towards the organization's innovative potential? Answers to such questions would free CI systems from being niche players in certain large-scale problem-solving initiatives.
Reflecting back from the potential of CI systems in today's organizations, we ask: what can be learned with respect to the design and implementation of future CI systems; and: how to break the silos, i.e., how to integrate them with latest computing resources such as big data and the data science toolkit?

I will present our generic temporal segmentation techniques that we have used for multiple applications. A new concept of cyber, physical, social contexts will be introduced, and how they translate in various domain applications of our research for analysing occupant behaviours, and for personalised intelligent assistants.
Finally, the data-driven models can be used as input for recommender systems for individuals and/or groups, and for optimising the routes and/or itineraries in multiple applications, from visitors in a shopping mall, to travelling parking inspectors and police officers.

Biography
Dirk Draheim is full professor of information systems and head of the Information System Group at Tallinn University of Technology. Dirk holds a Diploma in computer science from Technische Universität Berlin, a PhD from Freie Universität Berlin and a habilitation from the University of Mannheim. Until 2006, he worked as a Researcher at Freie Universität Berlin. From 2006-2008, he was area manager for database systems at the Software Competence Center Hagenberg, Austria. From 2008-2016 he was head of the data center of the University of Innsbruck and, in parallel, Adjunct Reader at the Faculty of Information Systems of the University of Mannheim. Dirk is co-author of the Springer book "Form-Oriented Analysis" and author of the Springer books "Business Process Technology", "Semantics of the Probabilistic Typed Lambda Calculus" and "Generalized Jeffrey Conditionalization". His research interest is the design and implementation of large-scale information systems.

Program Structure and Room allocation

Date	Time	ROOMS			
		FORUM 12	FORUM 13/14	FORUM 15	FORUM 16
Mon 2.12.2019	09:30	Welcome & Conference Opening			
	09:45	Keynote Talk 1			
	11:00	Keynote Talk 2			
	13:00	MoMM S.1A Context-aware Computing	iiWAS S.1A Classification Techniques & Applications	iiWAS S.1B Encryption and Security	iiWAS S.1C Machine Learning
	15:30	MoMM S.1B Mobile Sensing and Services	iiWAS S.1D Social Network Analysis	iiWAS S1E Social Network Analysis 2	iiWAS S.1F Text/Graph Analysis and Mining
Tue 3.12.2019	09:00	Keynote Talk 3			
	09:45	Panel			
	11:00	MoMM S.2A Wearables and Motion Sensing	iiWAS S.2A Machine Learning 3	iiWAS S.2B Social Network Analysis 3	iiWAS S.2C Trust and Privacy
	13:30	MoMM S.2B Mobile Biometrics and Security	iiWAS S.2E Semantic Web and Linked Data	iiWAS S.2A Database Systems	iiWAS S.2F Smart Environments and IoT
	16:00	MoMM S.2C Intelligent Agents	iiWAS S.2G Data and Data Mining	iiWAS S.2H Semantic Web & Linked Data 2	iiWAS S.2i HCI and Crowdfunding
Wed 4.12.2019	09:00	MoMM S.3A Multimedia and Computer	iiWAS S.3A Web Services and Applications	iiWAS S.3B Information Retrieval and Management	iiWAS S.3C Knowledge Representation
	11:00	MoMM S.3B Augmented Reality and Gamification	iiWAS S.3D IS Applications	iiWAS S.3E Machine Learning 2	S.3F SHOW
	13:30	MoMM S.3C IoT Systems	iiWAS S.3G Social Network Analysis 3	S.3H SHOW	S.3I SHOW
	16:00	MoMM S.3D Smart and Efficient Computing	iiWAS S.3J Hardware and Systems	S.3K SHOW	S.3L SHOW
	17:30	Closing			

Program Structure and Room allocation

All lecture rooms are located on the ground floor Holiday Inn Munich City Center.

Conference Program & Schedule

Sunday, 1st December 2019

TIME	
17.00-19.00	**Registration** Foyer Hallsaal (Ballroom), Ground Floor. Holiday Inn Munich – City Center

Monday, 2nd December 2019

TIME		
08.00	**Registration**	
09.30	**Welcome and Conference Opening**	Forum Room 12
09.45	**Keynote Talk I** ***Context and Behaviour Modelling of User and Urban Activities*** **Flora Salim,** CIDDA - RMIT University, Melbourne, Australia	Forum Room 12
10.30	Coffee Break @Forum 9 -11, Ground Floor	
11.00	**Keynote Talk II** ***Making Humans Smarter Than Artificial Intelligence:*** ***Digital Technologies to Amplify Human Perception and Cognition*** **Albrecht Schmidt,** LMU Munich, Germany	Forum Room 12
11.45	Group Photo Session	Foyer Ballsaal
12.00	Lunch @Restaurant Ground Floor	
13.00	Papers Presentations	

Session iiWAS – IA Forum room 13/14
Classification Techniques and Applications
Session Chair : Alexander Stenzer

Tourism application with CNN-Based Classification specialized for cultural information
Takuma Hirotsu, Masaharu Hirota, Tetsuya Araki, Masaki Endo, Hiroshi Ishikawa
Japan

Fake News Classification Based on Subjective Language
Caio Libanio Melo Jeronimo, Leandro Balby Marinho, Claudio E. C. Campelo, Adriano Veloso, Allan Sales da Costa Melo
Brazil

Building Classifier Models for on-off Javanese Character Recognition
Lucia D. Krisnawati, Aditya W. Mahastama
Indonesia

Movie Genres Classification using Collaborative Filtering
Raji Ghawi, Jürgen Pfeffer
Germany

13.00 Papers Presentations

Session iiWAS – IB Forum room 16
Encryption and Security
Session Chair : Chiemi Watanabe

Secure Naïve Bayes Classification Protocol over Encrypted Data Using Fully Homomorphic Encryption
Yoshiko Yasumura, Yu Ishimaki, Hayato Yamana
Japan

Outsourced Private Set Union on Multi-Attribute Datasets for Search Protocol using Fully Homomorphic Encryption
Rumi Shakya, Yoshiko Yasumura, Suzuki Takuya, Yu Ishimaki, Hayato Yamana
Japan

A Privacy-Preserving Query System using Fully Homomorphic Encryption with Real-World Implementation for Medicine-Side Effect Search
Yusheng Jiang, Tamotsu Noguchi, Nobuyuki Kanno, Yoshiko Yasumura, Takuya Suzuki, Yu Ishimaki, Hayato Yamana
Japan

Ontology-Based Model for Automotive Security Verification and Validation
Abdelkader Magdy Shaaban, Christoph Schmittner, Thomas Gruber, Abd El
Baith Mohamed, Gerald Quirchmayr, Erich Schikuta
Austria

Session iiWAS – IC Forum room 15
Machine Learning
Session Chair : Akiyo Nadamoto

A Hybrid Machine Learning Approach for Improving Mortality Risk Prediction on Imbalanced Data
Araek Tashkandi, Lena Wies
Germany

Analyzing and Predicting the Popularity of Online Contents
Minh-Tri Nguyen, Takuma Nakajima, Masato Yoshimi, Nam Thoai
Viet Nam, Japan

Flexible Community Search Algorithm on Attributed Graphs
Shohei Matsugu, Hiroaki Shiokawa, Hiroyuki Kitagawa
Japan

Fast RankClus Algorithm via Dynamic Rank Score Tracking on Bi-type Information Networks
Kotaro Yamazaki, Shohei Matsugu, Hiroaki Shiokawa, Hiroyuki
Japan

13.00 | Papers Presentations

Session MoMM – IA Forum room 12
Context-aware Computing
Session Chair : Piotr Wawryka

Using Personalisation to improve User Experience in Public Display Systems with Mobile Interaction
Rui Neves Madeira, Pedro Santos, Nuno Correia
Portugal

Context-Aware Smart Energy Recommender (CASER)
Paras Sitoula, Dwi Rahayu, Pari Delir Haghighi, Sarah Goodwin, Chris Ling
Australia

DCAT: A Deep Context-Aware Trust Prediction Approach for Online Social Networks
Seyed Mohssen Ghafari, Aditya Joshi, Amin Beheshti, Cecile Paris, Shahpar Yakhchi, Mehmet Orgun
Australia

Towards Context-aware Social Behavioral analytics
Amin Beheshti, Vahid Moraveji Hashemi, Shahpar Yakhchi
Australia

15.00 | Coffee Break @Forum 9 -11, Ground Floor

15.30 | Papers Presentations

Session iiWAS – ID Forum room 13/14
Social Network Analysis
Session Chair : Masaharu Hirota

Towards Linked Data for Wikidata Revisions and Twitter Trending Hashtags
Paula Dooley, Bojan Božić
Ireland

Enabling Social Information Exchange via Dynamically Robust Annotations
Vishwajeet Pattanaik, Shweta Suran, Dirk Draheim
Estonia

Blockchain as Middleware for Exchange of Resources: The Cross-Company Loyalty Rewarding System Use Case
Kushal Soni, Olga De Troyer
Belgium

Session iiWAS – IE Forum room 16
Social Network Analysis 2
Session Chair : Hiroaki Ohshima

A Simulation Model Demonstrating the Impact of Social Aspects on Social Internet of Things
Kashif Zia, Arshad Muhammad, Sanad Al Maskari, Dinesh Saini, Umar Farooq
Oman, Pakistan

15.30 Papers Presentations

Forum room 16

Theoretical approach to discover mutual friendships from social graph networks
Carson Leung, Fan Jiang, Sehaj Pal Singh, Alfredo Cuzzocrea
Canada, Italy

Analysis of High-value Reviews based on Sentiment
Rinji Suzuki, Kazuhiro Akiyama, Tadahiko Kumamoto, Akiyo Nadamoto
Japan

Forum room 15

Session iiWAS – IF
Text/Graph Analysis and Mining
Session Chair : Lucia Dwi Krisnawati

Optimal Selection of Training Courses for Unemployed People based on Stable Marriage Model
Jorge Martinez-Gil, Bernhard Freudenthaler
Austria

A Method to Estimate Entity Performance from Mentions to Related Entities in Texts on the Web
Vanderson Sampaio, Renato Fileto, Douglas Mace
Brazil

Pro-Eating Disorders and Pro-Recovery Communities on Reddit: Text and Network Comparative Analyses
Yousra Fettach, Lamia Benhiba
Morocco

Named entity recognition using point prediction and active learning
Koga Kobayashi, Kei Wakabayashi
Japan

Session MoMM – IB
Mobile Sensing and Services
Session Chair : Yusuke Gotoh

Forum room 12

Active Safety for Individual and Connected Vehicles using Mobile Phone Only
Mateusz Paciorek, Piotr Wawryka, Adrian Kłusek, Przemysław Kalawski, Michał Kosowski, Andrzej Piechowicz, Julia Plewa, Marek Powroźnik, Michał Śledź, Aleksander Byrski, Marcin Kurdziel, Wojciech Turek
Poland

A Hierarchical Classifier for Detecting Metro-Journey Activities in Data Sampled at Low Frequency
Ankita Dewan, Venkata M. V. Gunturi, Vinayak Naik, Kartik Vishwakarma and Shrehal Bohra
India

15.30	Papers Presentations	
	Scalable and Secure Mobile Sensor Cloud Architecture for the Delivery of Surveillance Applications as A Service Mohamed Jacem Guezguez, Slim Rekhis *Tunisia*	Forum room 12
	Continuous indoor/outdoor pathway display algorithm for pedestrian navigation service Wiem Fekih Hassen, Lionel Brunie, Asma Nechba, Harald Kosch *France, Germany*	
	A Division-based Broadcasting System Considering Dynamic Updates of Delivery Schedule Syunpei Kanamoto, Yusuke Gotoh *Japan*	
18.00 - 20.00	Welcome Reception at LMU Lichthof, the main building LMU Geschwister-Scholl-Platz 1. 80539 München	

Tuesday, 3rd December 2019

TIME		
08.30	Registration	
09.00	**Keynote Talk III** **Collective Intelligence Systems from an Organizational Perspective** **Dirk Draheim**, Tallinn University of Technology, Estonia	Forum Room 12
09.45	**Panel** **Big Data Management and Analytics in Intelligent Smart Environments: State-of-the-Art Analysis and Future Research Directions** **Alfredo Cuzzocrea,** ICAR-CNR and University of Calabria, Italy	
10.30	Coffee Break @Forum 9 -11, Ground Floor	
11.00	Papers Presentations	
	Session iiWAS – IIA **Machine Learning 3** Session Chair : Chinmaya Kumar Dehury	Forum room 13/14
	Film Genre Prediction Based on Film Content and Screenplay Structure Yusuke Nakano, Hiroaki Ohshima, Yusuke Yamamoto *Japan*	

11.00 Papers Presentations

Forum room 13/14

Personality Estimation using Demographic Data in a Personality-based Recommender System: A Proposal
Iman Paryudi, Ahmad Ashari, A Min Tjoa
Indonesia, Austria

Fatten Features and Drop Wastes: Finding Repeaters Reviews by Feature Generation and Feature Selection
Naoki Muramoto, Hiromi Shiraga, Kilho Shin, Hiroaki Ohshima
Japan

Effects of Mining Parameters on the Performance of the Sequence Pattern Variants Analyzing Method Applied to Electronic Medical Record Systems
Hieu Hanh Le, Tasuhiro Yamada, Yuichi Honda, Masaaki Kayahara, Muneo Kushima, Kenji Araki, Haruo Yokota
Japan

Session iiWAS – IIB
Social Network Analysis 3
Session Chair : Ngurah Agus Sanjaya ER

Forum room 16

Events Insights Extraction from Twitter Using LDA and Day-Hashtag Pooling
Muhammad Haseeb Ur Rehman Khan, Kei Wakabayashi, Satoshi Fukuyama
Japan

Modelling Emotion Dynamics on Twitter via Hidden Markov Mode
Debashis Naskar, Miguel Rebollo, Eva Onaindia, Subhashis Das
Spain, Italy

A Crawling Method with No Parameters for Geo-Social Data based on Road Maps
Sou Ijima, Masaharu Hirota, Shohei Yokoyama
Japan

Detection of Behavioral Facilitation information in Disaster Situation
Yoshiki Yoneda, Yu Suzuki, Akiyo Nadamoto
Japan

Session iiWAS – IIC
Trust and Privacy
Session Chair : Dirk Draheim

Forum room 15

Extending Linked USDL by Trust Assertion
Wolfgang Bauer, Natalia Kryvinska, Jürgen Dorn, Owen Sacco
Slovakia, Austria, Malta

URL-based Phishing Detection using the Entropy of Non-Alphanumeric Characters
Eint Sandi Aung, Hayato Yamana
Japan

11.00	Papers Presentations

Exploiting Blockchain and Smart Contracts for Data Exploration as a Service — Forum room 15
Ada Bagozi, Devis Bianchini, Valeria De Antonellis, Massimiliano Garda, Michele Melchiori
Italy

A Privacy-Preserving Similarity Search Scheme over Encrypted Word Embeddings
Daisuke Aritomo, Chiemi Watanabe, Masaki Matsubara, Atsuyuki Morishima
Japan

Session MoMM – IIA
Wearables and Motion Sensing — Forum room 12
Session Chair : Daniel Hintze

Tennis Stroke Classification: Comparing Wrist and Racket as IMU Sensor Position
Christopher Ebner, Rainhard Dieter Findling
Austria

Haptic Feedback Method using Deformation of Clothing
Kentaro Ueda, Tsutomu Terada, Masahiko Tsukamoto
Japan

Activity Recognition and Stress Detection via Wristband
Chun Yiu Wong, Jun Wang, Yujun Fu, Hong Va Leong, Grace Ngai
Hong Kong

An Auditory Feedback System to Improve the Foot Pressure Balance for Runners
Ayumi Ohnishi, Isao Nishiyama, Tsutomu Terada, Masahiko Tsukamoto
Japan

12.30	Lunch @ the Restaurant Ground Floor
13.30	Papers Presentations

Session iiWAS – IID
Semantic Web and Linked Data — Forum room 13.14
Jakub Klímek

PLDSD: Personalized Linked Data Semantic Distance for LOD-Based Recommender Systems
Gabriela da Silva, Frederico Durão, Miriam Capretz
Brazil, Canada

Knowledge Graph of University Campus Issues and Application of Completion Methods
Yuto Tsukagoshi, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga
Japan

13.30 Papers Presentations

Uniform Access to Multiform Data Lakes using Semantic Technologies
Mohamed Nadjib Mami, Damien Graux, Simon Scerri, Sören Auer, Jens Lehmann, Hajira Jabeen
Germany

Towards Knowledge Graph Construction using Semantic Data Mining
Dina Sharafeldeen, Alsayed Algergawy, Birgitta König-Ries
Germany

Session iiWAS – IIE
Database Systems
Session Chair : Lamia Benhiba

Forum room 16

On efficiently storing huge property graphs in relational database management systems
Matthias Schmid
Germany

A Hybrid Partitioning Strategy for NewSQL Databases: The VoltDB Case
Geomar Schreiner, Denio Duarte, Guilherme Dal Bianco, Ronaldo Mello
Brazil

NewSQL Through the Looking Glass
Geomar Schreiner, Ronan Knob, Denio Duarte, Patricia Vilain, Ronaldo Mello
Brazil

Structural Transition Analysis of Dynamic Network Based on Roles of Adding Edges
Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh
Japan

Session iiWAS – IIF
Smart Environments and IoT
Session Chair : Tatsuo Nakajima

Forum room 15

Weighted Load Balancing in Distributed Hash Tables
Robin Lösch, Jan Schmidt, Nils Gentschen Felde
Germany

Open Routed Energy Distribution Network based on a Concept of Energy Routers in Smart Grid
Alexander Dudko, Tatiana Endrjukaite, Leon L. Roose
Japan, Latvia, USA

Microbiological Water Quality Test Results Extraction from Mobile Photographs
Jifang Xing, Ruixi Zhang, Remmy Zen, Ngurah Agus Sanjaya Er, Laure Sioné, Ismail Khalil, Stéphane Bressan
Singapore, Indonesia, UK, Austria

11.30 Papers Presentations

Building Extraction from Google Earth Images
Jifang Xing, Ruixi Zhang, Remmy Zen, Dewa Made Sri Arsa, Ismail Khalil, *Stéphane Bressan*
Singapore, Indonesia, Austria

Session MoMM – IIB Forum room 12
Mobile Biometrics and Security
Session Chair : Ayumi Ohnishi

Hide my Gaze with EOG! Towards Closed-Eye Gaze Gesture Passwords that Resist Observation-Attacks with Electrooculography in Smart Glasses
Rainhard Dieter Findling, Tahmid Quddus, Stephan Sigg
Finland

CORMORANT: On Implementing Risk-Aware Multi-Modal Biometric Cross-Device Authentication for Android
Daniel Hintze, Matthias Füller, Sebastian Scholz, Rainhard Dieter Findling, Muhammad Muaaz, Philipp Kapfer, Wilhelm Nüßer, Rene Mayrhofer
Austria, Finland, Germany

Security solution methods in the Vehicular Ad-Hoc Networks
Krzysztof Stępień, Aneta Poniszewska-Maranda
Poland

Free-Form Gaze Passwords from Cameras Embedded in Smart Glasses
Eira Friström, Elias Lius, Niki Ulmanen, Paavo Hietala, Pauliina Kärkkäinen, Tommi Mäkinen, Stephan Sigg, Rainhard Dieter Findling
Finland

15.30 Coffee Break @Forum 9 -11, Ground Floor

16.00 Papers Presentations

Session iiWAS – IIG Forum room 13/14
Data and Data Mining
Session Chair : Yoshihisa Udagawa

Improving Findability of Open Data Beyond Data Catalogs
Tomáš Skopal, Jakub Klímek, Martin Nečaský
Czech Republic

An Application of Distributed Data Mining to Identify Data Quality Problems
Eshref Januzaj, Visar Januzaj, Peter Mandl
Germany

ToT for CSV: accessing open data CSV files through SQL
Yasushi Doi, Motomichi Toyama
Japan

Generation of Test Cases for Testing SuperSQL
Amulya Bathini, Kento Goto, Motomichi Toyama
Japan

16.00 Papers Presentations

Session iiWAS – IIH Forum room 16
Semantic Web & Linked Data 2
Session Chair : Kei Wakabayashi

Query Relaxation using Spreading-Activation and SKOS-Ontologies
Alexander Stenzer
Germany

Reverse-Transliteration of Hebrew script for Entity Disambiguation
Aaron Christianson, Maral Dadvar, Kai Eckert
Germany

Social Media Copyright Management using Semantic Web and Blockchain
Roberto Garcia, Rosa Gil
Spain

Session iiWAS – II I Forum room 15
HCI and Crowdfunding
Session Chair : Yujun Fu

Gamifying Human Behavior in Urban Crowdsourcing for a Sustainable Smart City
Risa Kimura, Tatsuo Nakajima
Japan

Query Recommendation to Draw a Laugh from Web Searcher
Hiroo Umeda, Yusuke Yamamoto
Japan

What Independent Game Developers Expect from Recommender Systems
Marta Kholodylo, Christine Strauss
Austria

The Impact of Updates in Social Crowd Projects: Insights from a German Equity Crowdfunding Platform
Regina Cernicka, Andreas Mladenow, Christine Strauss
Austria

Session MoMM – IIC Forum room 12
Intelligent Agents
Session Chair : Tsutomu Terada

Blockchain e-voting system with the use of intelligent agent approach
Michał Pawlak, Aneta Poniszewska-Maranda
Poland

A Reinforcement Learning and Synthetic Data Approach to Mobile Notification Management
Rowan Sutton, Kieran Fraser, Owen Conlan
Ireland

16.00	Papers Presentations

Be in/Be out model for intelligent transport in SmartCity approach — Forum room 12
Bartosz Wieczorek, Aneta Poniszewska-Maranda
Poland

18.30-22.30	**Conference Gala Dinner & Best Paper Awards** a traditional Bavarian Dinner at Augustiner Keller München, Arnulfstr. 52, 80335 München

Wednesday, 4th December 2019

09.00	Papers Presentations

Session iiWAS – IIIA — Forum room 13/14
Web Services and Applications
Session Chair : Lukito Edi Nugroho

Building a Web-Based Federated Toolchain: Lessons Learned from a Four-Year Industrial Project
Damir Nešić, Jad El-Khoury, Jonas Westman, Mattias Nyberg
Sweden

Personalized Service Delivery using Reinforcement Learning in Fog and Cloud Environment
Chinmaya Kumar Dehury, Satish Narayana Srirama
Estonia

A Web Serivce Architecture for Social Micro-Learning
Bernhard Göschlberger, Gabriele Anderst-Kotsis
Austria

Computing Ranges for Temporal Parameters of Composed Web Services
Marco Franceschetti, Johann Eder
Austria

Session iiWAS – IIIB — Forum room 16
Information Retrieval and Management
Session Chair : Brahim Ouhbi

Improving Precision in IR Considering Dynamic Environments
Claudio Gutiérrez-Soto, Arturo Curiel
Chile, Mexico

Target-Topic Aware Doc2Vec for Short Sentence Retrieval from User Generated Content
Kosuke Kurihara, Yoshiyuki Shoji, Sumio Fujita, Martin J.Dürst
Japan

09.00 Papers Presentations

SIMPATICO 3D Mobile for Diagnostic Procedures
Ester Zumpano, Pasquale Iaquinta, Francesco Dattola, Luciano Caroprese, Giuseppe Tradigo, Pierangelo Veltri, Eugenio Vocaturo
Italy

Session iiWAS – IIIC Forum room 15
Knowledge Representation
Session Chair: Ismail Khalil

Genereting Anthropomorphism of Subject and Verb by Transformation Matrix
Katsurou Takahashi, Kilho Shin, Hiroaki Ohshima
Japan

Fully Informed Vulnerable Road Users - Simpler, Maybe Better
Bruno Fernandes, Henrique Vicente, Jorge Ribeiro, Antönio Capita, Cesar Analide, Jose Neves
Portugal

HyDRS-WoT: Community-oriented Hybrid Disaster Response System using Web of Things
Youna Jung, Aaron Causey
USA

Session MoMM – IIIA Forum room 12
Multimedia and Computer Vision
Session Chair : Shahpar Yakhchi

Large-Scale Semantic Concept Detection Based on Visual Contents
Mohamed Hamroun, Sonia Lajmi, Henri Nicolas, Ikram Amous
Tunisia, France

Automatic Vehicle Identification Through Visual Features
Imran Shafiq Ahmad, Boubakeur Boufama
Canada

An Approach for CCTV Contents Filtering Based on Contextual Enrichment via Spatial and Temporal Metadata
Franck Jeveme Panta, André Péninou, Florence Sedes
France

10.30 Coffee Break @Forum 9 -11, Ground Floor

11.00 Papers Presentations

Session iiWAS – IIID Forum room 13/14
IS Applications
Session Chair : Yusuke Yamamoto

Prototyping and Preliminary Evaluation of Mind Monitoring Service for Elderly People at Home
Chisaki Miura, Haruhisa Maeda, Masahide Nakamura, Kiyoshi Yasuda
Japan

A Museum Information System for Sustaining and Analyzing National Cultural Expressions
Bea Therese Santos, Jacob Tolentino, Danielle Nicole Aquino, Riel Malibiran, Christine Diane Ramos, Charibeth Cheng, Courtney Ngo
Philippines

A Context-Aware Adaptive Tourist Recommendation System
Lukito Edi Nugroho, Rico Yudha Saputra, Vivin Mahat Putri, Yohandes Efindo
Indonesia

Session iiWAS – IIIE Forum room 16
Machine Learning 2
Session Chair : Ester Zumpano

Accent neutralization for speech recognition of non-native speakers
Kacper Radzikowski, Mateusz Forc, Le Wang, Osamu Yoshie, Robert Nowak
Japan, Poland

Mining Stock Price Changes for Profitable Trade Using Candlestick Chart Patterns
Yoshihisa Udagawa
Japan

Estimation Method of L2 Learners' Second Language Ability by using Features in Conversation
Xinnan Chen, Muhammad Haseeb Ur Rehman Khan, Wakabayashi Kei
Japan

Session SHOW – IIIF Forum room 15
Session Chair: Gabriele Anderst-Kotsis

Building Blocks of Negotiating Agents for Healthcare Data
Svetlana Boudko, Wolfgang Leister
Norway

Towards a social learning environment
Hajar Zankadi, Imane Hilal, Najima Daoudi, Abdellah Idrissi
Morocco

Data Source Selection in Big Data Context
Hicham Moad Safhi, Bouchra Frikh, Brahim Ouhbi
Morocco

11.00	Papers Presentations

Session MoMM – IIIB Forum room 12
Augmented Reality and Gamification
Session Chair: Marco Franceschetti

An Interaction Design Model for Information Visualization in Immersive Augmented Reality platform
Shafaq Irshad, Dayang Awang, Suzaih Sulaiman
Malaysia

Evaluation of Effect on Walking Behavior by Seeing Augmented Reality Objects
Yamato Sakuragi, Naoya Isoyama, Tsutomu Terada, Masahiko Tsukamoto
Japan

12.30	Lunch @ the Restaurant Ground Floor

13.30	Papers Presentations

Session iiWAS – IIIG Forum room 13/14
Social Network Analysis 3
Session Chair: Khashif Zia

An Analysis of Influence of Emoticons on Affective Impressions Feeling from Tweets
Koji Nakahira, Tadahiko Kumamoto
Japan

Method for Computing Emotions of Tweets with an Emoticon
Chengzhi Jiang, Tadahiko Kumamoto
Japan

A Study on Characterizing the Ecosystem of Monetizing Video Spams on YouTube
Ashutosh Tripathi, Kusum Kumari Bharti, Mohona Ghosh
India

Session SH0W – IIIH Forum room 16
Session Chair: Gabriele Anderst-Kotsis

Named Entity Recognition for Biomedical Patent Text using Bi-LSTM Variants
Farag Saad
Germany

Robust and Scalable Pipeline for the Real-time Processing and Analysis of Massive 3D Spatial Streams
Salman Ahmed Shaikh, Jun Lee, Akiyoshi Matono,Kyoung-Sook Kim
Japan

Security Issues in Mobile Healthcare Applications
Youna Jung
USA

BM25-AH: Enhanced BM25 Algorithm for Domain-Specific Search Engine
Youna Jung
USA

13.30 Papers Presentations

Session SHOW– III I Forum room 15
Session Chair: Dirk Draheim

Study of machine learning methods for customer churn prediction in telecommunication company
Anna Śniegula, Aneta Poniszewska-Marańda
Poland

Facilitation of Domain-Specific Data Models Design using Semantic Web Technologies for Manufacturing
Vaclav Jirkovsky, Ondrej Sebek
Czech Republic

Session MoMM – IIIC Forum room 12
IoT Systems
Session Chair : Imran Ahmad

Gateways to Heaven: Observations and Predictions on the Software Architecture of IoT Gateways
Antero Taivalsaari, Tommi Mikkonen
Finland

Mitigating Messaging and Processing Load in IoT Environments Managed by Blockchain
Johannes Mittendorfer, Karin Anna Hummel
Austria

A Holistic Approach for Heating and Ventilation Control in EEBs
Abdelhak Kharbouch, Mohamed Bakhouya, Anas El Maakoul, Driss El Ouadghiri
Morocco

15.30 Lunch @ the Restaurant Ground Floor

16.00 Papers Presentations

Session iiWAS – IIIJ Forum room 13/14
Hardware and Systems
Session Chair : Alfreddo Cuzzocrea

Development of IoT Monitoring Device and Prediction of Daily Life Behavior
Rabin Maharjan, Koichi Shiraishi, Takehiro Yamamoto, Yusuke Yamamoto, Hiroaki Ohshima
Japan

A Design of an Anti-Phishing Training System Collaborated with Multiple Organization
Masayuki Higashino
Japan

16.00 | Papers Presentations

Session SHOW – IIIK — Forum room 16
Session Chair: Ismail Khalil

A Secure Used Car Trading System based on Blockchain Technology
Cho Cho Htet, May Htet
Myanmar

Predicting the Category of Fire Department Operations
Kevin Pirklbauer, Rainhard Dieter Findling
Finland, Austria

Rule-Based Inquiry Service to Elderly at Home for Efficient Mind Sensing
Haruhisa Maeda, Sachio Saiki, Masahide Nakamura, Kiyoshi Yasuda
Japan

Content Based Fake News Detection Using N-Gram Models
Hnin Ei Wynne, Zar Zar Wint
Myanmar

Session SHOW – IIIL — Forum room 15
Session Chair: Dirk Draheim

Conversational AI for Corporate e-Learning
Bernhard Göschlberger, Christoph Brandstetter
Austria

Combining a Gamified Civic Engagement Platform with a Digital Game in a Loosely Way to Increase Retention
Renny Lindberg, Jan Maushagen, Olga De Troyer
Belgium

Your body signals expose your fall
Eugene Yujun Fu, CheukYin Wong, Katie TY Lau, Hong Va Leong, Grace Ngai
Hong Kong

Session MoMM – IIID — Forum room 12
Smart and Efficient Computing
Session Chair : Pari Delir Haghighi

An Adaptive Recursive Method of ESS for Peak Shaving in Zero Energy Community
Hong-Soon Nam
South Korea

VComputeLib: Enabling Cross-Platform GPGPU on Mobile and Embedded GPUs
Nadjib Mammeri, Ben Juurlink
Germany

Fun with AEDs: Examining the Effects of a Gamified Mobile Crowdsourcing Application
Anran Hao, Chei Sian Lee, Chee Wah Wesley Tan
Singapore

17.30 | Forum Room 12, Ground Floor
Closing

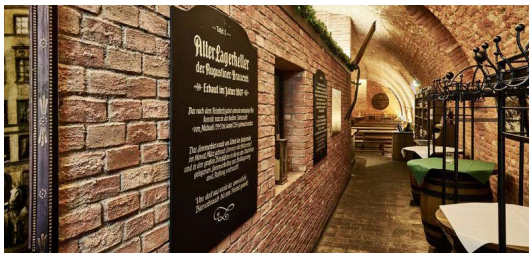
Welcome Reception

Welcome Reception will be held on the first day evening at the LMU Lichthof, - the main Building of LMU, Geschwister-Scholl-Platz 1. 80539 München, hosted by @WAS organization and Ludwig-Maximilians-University (LMU) Munich. There will be drinks and light finger-food to serve while guests enjoy a harmonious work of art between the aesthetics and architecture as well as the historical significance of the atrium with the impressive light installations.

To get to the LMU Lichthof from Holiday Inn Munich City Center, you take one of the S-bahn (train) lines 1, 2, 4, 6, 7, or 8 from “Rosenheimer Platz” and stop at “Marienplatz” station. From there you can:
- change to subway line U3 and exit at station “Universität” or
- walk from Marienplatz station to LMU main building, it takes about 17 minutes walking.

Conference Gala Dinner

On the 2nd day evening, A Gala Dinner with a traditional Bavarian menu will be held at at Augustiner Keller München, Arnulfstr. 52, 80335 München.
The Lagerkeller is reserved for iiWAS/MoMM 2019 Conference Dinner exclusively and partcipants will have the opportunity to taste Bavarian beer as well as hearty local food.



To get to Augustiner Keller München from Holiday Inn Munich City Center: Take one of S-bahn (train) lines 1, 2, 3, 4, 6, 7, or 8 from “Rosenheimer Platz” to stop at “Hackerbrücke” station, from there the venue is 5 minutes walking distance.

Conference History

The International organization on Information Integration and Web-based Applications & Services (@WAS) was founded in 2003 given the success of the iiWAS series of workshops (1999, 2000) and conferences (2001-on) and the prospect of continuing the organization of these events in the forthcoming years which have encouraged some to propose the creation of a non-profit International Organization for Information Integration and Web-based Applications & Services (@WAS).

In 2003, @WAS launched another series of conferences called International Conferences on Advances in Mobile Computing and Multimedia (MoMM) held in Jakarta in conjunction with iiWAS2003 conference and was officially opened by the Minister of Communication and Information Technology of the Republic of Indonesia at that time.
MoMM aims to provide a discussion forum for the exchange of ideas and Information by researchers, students, and professionals on the issues and challenges brought by the emerging wireless technologies for mobile and multimedia applications and services.

Conference History iiWAS

	2003	2004	2005	2006	2007	2008	2009
Location	Jakarta, Indonesia	Jakarta, Indonesia	Kuala Lumpur, Malaysia	Yogyakarta, Indonesia	Jakarta, Indonesia	Linz, Austria	Kuala Lumpur, Malaysia
Dates	15-17 Sep	27-29 Sep	19-21 Sep	4-6 Dec	3-5 Dec	24-26 Nov	14-16 Dec
No. of Participants	57	106	129	127	135	110	146
No. of Accepted Papers	36	91	119	53	32	52	102
Acceptance rate	46%	40%	33%	28%	27%	30%	23%
Keynote Talks	4	4	3	3	3	3	5
Tutorials	4	5	3	2	1	3	5
Workshops	-	-	-	4	3	2	3
Panel	1	-	-	-	1	-	
Proceedings	OCG	OCG	OCG	OCG	OCG	ACM & OCG	ACM & OCG
Journals Special Issues	-	4	5	4	4	4	6

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Location	Paris, France	Ho Chi Minh City, Vietnam	Bali, Indonesia	Vienna, Austria	Hanoi, Vietnam	Brussel, Belgium	Singapore	Salzburg, Austria	Yogyakarta, Indonesia
Dates	8-10 Nov	5-7 Dec	3-5 Dec	2-4 Dec	4-6 Dec	11-13 Dec	28-30 Nov	4-6 Dec	19-21 Nov
No. of Participants	119	108	112	139	95	117	75	103	76
No. of Accepted Papers	82	93	34	43	37	58	41	90	65
Acceptance rate	38%	23%	30%	31%	39%	40%	45%	40%	40%
Keynote Talks	4	4	6	4	4	4	2	3	3
Tutorials	4	0	0	0	0	0	0	0	0
Workshops	3	0	0	0	0	0	0	1	1
Panel	-	1	1	0	1	0	1	1	0
Proceedings	ACM & OCG	ACM	ACM	ACM	ACM	ACM	ACM	ACM	ACM
Journals Special Issues	6	4	3	4	4	4	2	2	2

Conference History MoMM

	2003	2004	2005	2006	2007	2008	2009
Location	Jakarta, Indonesia	Bali, Indonesia	Kuala Lumpur, Malaysia	Yogyakarta, Indonesia	Jakarta, Indonesia	Linz, Austria	Kuala Lumpur, Malaysia
Dates	15-17 Sep	22-24 Sep	19-21 Sep	4-6 Dec	3-5 Dec	24-26 Nov	14-16 Dec
No. of Participants	28	68	45	73	82	73	102
No. of Accepted Papers	15	52	36	33	82	37	64
Acceptance rate	53%	51%	40%	39%	28%	32%	28%
Keynote Talks	-	2	3	2	2	3	5
Tutorials	-	2	1	1	1	3	5
Workshops	-	-	-	3	3	4	5
Panel	-	-	-	-	1	-	1
Proceedings	OCG	OCG	OCG	OCG	OCG	ACM & OCG	ACM & OCG
Journals Special Issues	1	4	4	2	3	4	6

	2010	2011	2012	2013	2014	2015	2016
Location	Paris, France	Ho Chi Minh City, Vietnam	Bali, Indonesia	Vienna, Austria	Kaohsiung, Taiwan	Brussel, Belgium	Singapore
Dates	8-10 Nov	5-7 Dec	3-5 Dec	2-4 Dec	8-10 Dec	11-13 Dec	28-30 Nov
No. of Participants	69	64	78	129	46	77	56
No. of Accepted Papers	39	58	27	39	38	36	25
Acceptance rate	33%	35%	35%	33%	45%	47%	45%
Keynote Talks	4	4	6	4		4	2
Tutorials	4	0	0	0		0	0
Workshops	3	0	0	0		0	0
Panel	-	1	1	0		0	1
Proceedings	ACM & OCG	ACM	ACM	ACM	ACM	ACM	ACM
Journals Special Issues	6	4	4	3	3	4	4

	2017	2018
Location	Salzburg, Austria	Yogyakarta, Indonesia
Dates	4-6 Dec	19-21 Nov
No. of Participants	42	22
No. of Accepted Papers	35	19
Acceptance rate	45%	40%
Keynote Talks	3	3
Tutorials	0	0
Workshops	1	1
Panel	1	0
Proceedings	ACM	ACM
Journals Special Issues	4	4

@WAS iiWAS 2019 MoMM 2019

December 2-4, 2019, Munich, Germany

CERTIFICATE

This Certificate is awarded to

Dr. Lucia Dwi Krisnawati
Duta Wacana Christian University
Indonesia

In recognition of the significant contribution to
The 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS2019)

as **Author**

Munich, 2-4 December 2019

Ismail Khalil
WAS Organization President

Gabriele Anderst-Kotsis
General Conference Chair

Building Classifier Models for on-off Javanese Character Recognition

Lucia D. Krisnawati
Universitas Kristen Duta Wacana
Jogjakarta, Indonesia
krisna@staff.ukdw.ac.id

Aditya W. Mahastama
Universitas Kristen Duta Wacana
Jogjakarta, Indonesia
mahas@staff.ukdw.ac.id

ABSTRACT

In this paper, we demostrated the building process of four classifier models as a part of an on-off character recognition system for Javanese characters. As Javanese character is no longer used in everyday writing and books, the dataset were collected by scanning the historical manuscripts and a reading lesson book. The rough dataset comprises 15.414 annotated characters and 633 classes. However, only 162 classes have sufficient data samples to be the training and testing one. Using this dataset, we measured the performance of four classifiers, namely k-NN, LDA, SVM, and Gaussian NB on the accuracy, micro-averaged precision, micro-averaged sensitivity and weighted-averaged precision and sensitivity metrices. The experiment shows that k-NN outperforms any other classifiers almost in most metrices, while SVM suffers the poorest performance. The research byproduct worth mentioning here is that it has identified 633 classes of distinct Javanese characters which comprise both common characters and compound characters found in modern Javanese writing as well as the archaic characters found in the literary works only.

CCS CONCEPTS

• **Applied computing → Optical character recognition**; *Document capture.*

KEYWORDS

Optical Character Recognition, Javanese Characters, Linear Discriminant Analysis, Support Vector Machine, k-NN

ACM Reference Format:
Lucia D. Krisnawati and Aditya W. Mahastama. 2019. Building Classifier Models for on-off Javanese Character Recognition . In *The 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS2019), December 2–4, 2019, Munich, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3366030.3366050

1 INTRODUCTION

Optical Character Recognition (OCR) has been applied in various fields such as in writer identification for forensic research [7], indexing and searching the archival documents and manuscripts [16], or in recognizing the vehicle lisence plate [2]. Massively, OCR has been applied in digitization projects of recent and early printed books for the sake of building e-library [19, 20]. The projects of preserving the content of historical manuscripts, books, or newspapers implement character recognition on various writing systems such as Latin alphabets [17], Tangut [13], Thai, Bangla and Latin [22], Arabic alphabets [10], and Javanese characters [6, 8, 15] as well.

As an active area of research, character recognition inclines to be a part of OCR building blocks. OCR is a field of study which deals with converting digital images of text into editable documents that can be processed, edited, searched, saved, or copied using a computer application [3]. To achieve this broad and complex task, the pipeline of OCR system is conventionally decomposed into the following sequences of processing: image acquisition, preprocessing, segmentation, feature extraction, character recognition and post-processing [5, 9]. The digital images of text are acquired by either scanning or photographing, while the preprocessing phase covers the noise reduction, binarization, and layout analysis [20]. The so-called character recognition very often falls into a task of classification which assigns an inputted character into a predefined class of characters. The post-processing phase deals with the correction of the recognition errors in order to increase the accuracy rate.

The main goal of the OCR research is to gain the high rate of recognition accuracy which remains a challenging problem. In OCR researches, this problem is addressed differently since each step of OCR pipeline contributes to the success of the recognition process. Additionaly, the text genre such as historical manuscripts or modern well printed texts present different contrainsts. The historical texts tend to have a greatly degraded image quality [16]. Therefore, some researches addressed the preprocessing problems such as noise removal and geometric correction [4], problem of text, line, and character segmentation [14, 17], finding a robust feature extraction method for historical handwritten texts as found in [22], the problem of recognition by word matching [19] or classification models [8], or correcting the recognition error in the post-processing step as in [1, 3].

In this study, we investigated the performance of several classifier algorithms to recognize Javanese characters. The use case of our OCR system is that it receives an input in the form of a text image online. In the experiment context, the text images were acquired by scanning text taken from *Serat Mangkunegaran IV*, which is a Javanese book written at the end of 19^{th} century. The system proceeds its task by gray scaling and binarizing the text image, segmenting line and characters, and extracting each of the inputted character features. The stream of character features become the

iiWAS2019, December 2–4, 2019, Munich, Germany

ACM ISBN 978-1-4503-7179-7/19/12…$15.00
https://doi.org/10.1145/3366030.3366050

input to the classifiers which should be able to predict to which class each of these characters belongs. The recognition outputs will be displayed online. The problem we faced is that we have very limited annotated character instances in each class of characters as training data. Therefore, we compared the performance of the following classification models: K-Nearest Neighbor (k-NN), Gaussian Naive Bayes (GNB), Linier Discriminant Analisys (LDA), and Support Vector machine (SVM).

The remaining part of this paper is organized as follows: Section II highlights the previous works related to this study, while Section III presents the brief overview on the Javanese characters. Section IV deals with the concepts of classification models applied in this study. Section V describes the pipeline of the proposed methods and its architecture. The strategy for building the data set, the experiment scenario and the evaluation on each classifier performance will be presented on Section VI. The last section deals with the conclusion of this study.

2 RELATED WORK

Based on the way a text is written, OCR researches focus on either printed or handwritten character recognition. This option is based on the more complex application which applies the OCR results. For example, digitizing projects for the sake of building e-library tend to focus more on the printed character recognition [20], while the projects on writer identification in historical texts incline to delve deeper on the handwritten one [7, 22].

Based on the way the character input is received, the OCR system could be categorized into an online character recognition (OLCR) or offline character recognition (OFCR). The significant difference between these systems is that OLCR processes temporal data using a list of points representing the track of writing [10], while OFCR processes an image as input which has been digitized, stored, and processed by the OCR system. This dichotomy gives no place to an OCR system accessible through Internet and receives text images as input online and delivers its output online but performs its recognition offline. Such system differs from the one that combines OLCR and OFCR as found in [10, 18]. From this point forward, we would like to address such system as an on-off character recognition system.

Related to OLCR and OFCR, Hamdani et. al. [10] combined the offline and online systems at the feature level. They applied beta-elliptic approach in capturing the spatio-temporal information in the feature extraction step for offline handwritten recognition process. The classifier was built using HMM-toolkit provided by Mathlab [10]. Surinta et. al. [22] built an offline handwritten character recognition which applied k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM) as classifiers and highlighted the use of local gradient feature descriptor on the feature extraction phase. Focusing more on the automated layout analysis, Reul, et al. [10] built an offline OCR system using OCRopus library for recognizing early printed books from 15^{th} century.

In recognizing characters, the most common approach used is classification with its various learning models. However, classification is not the only option for recognizing character as shown in [16, 19]. Instead of classification, they utilized word spotting and word matching by means of Dynamic Time Warping (DTW) which aligns and compares sets of features extracted from two images of words [19]. Having sparse data training, Liu [13] implemented Principle Component Analysis (PCA) based classifier to recognize Tanggut characters. Meanwhile, Namsyl and Konya [17] combined deep learning (CNN-LSTM models) with synthetic training data generation and data augmentation technique to build a segmentation-free OCR system. They claimed that their system architecture can be used to recognize printed, handwritten, and scene texts [17] as well.

Using classifiers for recognizing characters, most researches on OCR for Javanese characters tend to train their prototypes with the *incomplete* Javanese characters as dataset. Included in Abugida writing style, Javanese script has 20 basic consonants carrying inherent vowel 'a' called *Nglegena*. Other vowels and consonant cluster builders are written as diacritics which results in a compound character. Unfortunately, there has been no definite study stating the exact number of Javanese characters including the compound ones. A study in [24] assumed that the total number of Javanese characters is ca. 103 by excluding the compound characters. However, Unicode provides 91 code points including numbers and punctuation marks for Javanese characters.

In recognizing Javanese characters, Budhi and Adipranata [8] applied several ANN classifiers trained with 31 distinct characters along with 20 samples for each character class, while Dewa et. al. [6] applied CNN and Multilayer perceptron (MLP) to be tested with handwritten *Nglegena*. Widiarti and Wastu [24] developed an HMM-based classifier model for recognizing *Nglegena* too. As data set, they took 1000 characters consisting of 50 samples for each *Nglegena* character. Akin to Dewa et al., Karundeng et al. [11] made use handwritten *Nglegena* as much as 3379 samples to be trained and tested to their SVM-based classifier. In contrast to these Javanese OCR systems, this study made use of 162 classes of Javanese characters with 10 samples for each class. These classes comprise Nglegena, diacritics, numbers, punctuation marks, and some compound characters. In total, our dataset consists of 1620 characters.

3 JAVANESE CHARACTERS IN A NUTSHELL

Javanese is considered to be one of world's classical languages with literary tradition over a thousand years. Javanese has been written in Javanese script which is an Abugida type – a segmental writing system in which consonant-vowel sequences are written as a syllable unit [12]. In total, there are 20 basic consonant-vowel sequences as displayed in Figure 1. In Figure 1, the *Nglegena*, well knowkn also as *carakan*, are placed on the rows with the heading *aksara*, their consonant cluster builders are put on rows with the heading *Pasangan*. It displays their transliteration in Latin alphabet found under the heading *Nama*. The sign ©shows the position of *Carakan* and whether *Pasangan* should be written under *Carakan* functioning as diacritics or on the right side of *Carakan*.

Javanese texts are written from left to right and recognize no word boundary (*Scriptio Continua*). The presence of *pasangan* might indicate the first syllable of another word, yet it is not always the case. The main function of *Pasangan* is to start a succeeding syllable by making a *Carakan* preceeding it to be a consonant, i.e it nulifies the vowel of its preceeding syllable and thus make it a

Nama	Ha	Na	Ca	Ra	Ka
Aksara					
Pasangan					
Nama	Da	Ta	Sa	Wa	La
Aksara					
Pasangan					
Nama	Pa	Dha	Ja	Ya	Nya
Aksara					
Pasangan					
Nama	Ma	Ga	Ba	Tha	Nga
Aksara					
Pasangan					

Figure 1: The basic character sequences of Javanese (Nglegena) along with their consonant cluster builders (Pasangan)

Figure 2: An example of writing 3 words at the end of a setence in Javanese characters

closed syllable. Figure 2 highlights *Pasangan* 'ka' and 'ja' in red boxes, which introduce the first syllables of the following words, while *pasangan* 'ka' in 'kka' starts its successive syllable only. The consonant cluster builder within a syllable is written by a group of diacritics called *Sandangan* which is written either on the left, right, above or continuously attached under the *Carakan (Nglegena)*. Figure 3 gives examples on how to write some consonant cluster builders within a syllable. Other vocals such as *o, i, u, è, ê* are treated as diacritics, but they will be written as special compound characters when they are used to indicate vowel-initial syllables or to write loanwords.

Interistingly, Javanese writing system has morphographemic rules to mark the boundary of a free morpheme when it is followed by a bound morpheme (suffixes) in the form of a vowel-initial syllable. The rule requires to write the suffix using the coda (last consonant) of its preceeding syllable as its onset. This results in the presence of twin consonants. To make it clear, the second word in Figure 2 depicts a morphographemic rule which is used to mark the boundary of a suffix to its root word. Literary transliterated, the complex word *kaklempakaken* with the base word *klempak* – which is underlined – has to be written *kaklempakkaken* in which the double graphemes 'kk' indicate the presence of a suffix *aken*.

Figure 3: Some examples on characters functioning as consonant cluster builder within syllables and thus form inseparable compound characters

With such writing system, it remains a question left: how many distinctive characters the Javanese script has. To answer this question, we relied on the output of the segmentation process reported in [15]. The rationale is that there has been no study explicitly answering this question. The 91 blocks provided by unicode or a set of 103 characters mentioning in [24] list only character categories which are combinable to form compound characters. However many compound characters are inseparable as shown in Figure 3. The significance of having the number of distinctive characters is to obtain the number of classes to classify. Using 72 scaned pages which resulted in 19.112 segments of characters, Mahastama and Krisnawati [15] identified that there are 633 distinctive characters which include numbers and punctuations as well.

4 CLASSIFICATION MODELS

This section presents some classification models that have been implemented in this study. It starts with a brief description on k-Nearest Neighbors (k-NN), followed by Naive Bayes classifier and Linear Discriminant Analysis (LDA). The Support Vector Machine ends the discussion on the experimented classifier concepts.

4.1 K-Nearest Neigbors

Well known as a simple algorithm in Machine Learning (ML), k-NN is categorized as a nonparametic and instance-based learning method. It is nonparametic in the sense that it does not assume data distribution [5], the data points themselves become the parameters. As an instance-based method, k-NN's learning process is simply storing all instances of the training data. When a new instance or test data point is encountered, it will be matched to all instances in the training data.

In k-NN, the training phase is done by adding each sample instance to the list of *data samples*. Given an instance x_q, the k-NN classification algorithm will compute the distance of x_q to all data

samples and sort the distances in ascending order. Then, it discriminates k number of training instances, x_i, which are nearesst to x_q. The distance metric applied is the Minkowski distance, which is a generalization of both Euclidean and Manhattan distances. The Equation 1 shows how to compute the Minkowski distance which is commonly calculated with p being 1 or 2 . If p equals to 1, then it corresponds to the use of Manhattan distance, while p equals to 2 refers to the use of Euclidean one.

$$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} |P_i - Q_i|^p} \tag{1}$$

where P denotes to parameters, and Q to queries.

The class prediction of x_q is done by voting and the class of an instance with the maximum votes in k neigbors is taken as the prediction. The computation of a class prediction is presented in Equation 2. It is clear that the function $\hat{f}(x_q)$ is assigned the value of $f(x_i)$ where x_i is the training instance nearest to x_q with the maximum voting, v, where $\delta(a, b) = 1$ if $a = b$ and 0 otherwise.

$$\hat{f}(x_q) \leftarrow \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v, f(x_i)) \tag{2}$$

With such algorithm, one of k-NN drawbacks is that it requires high memory for storing and load all training data in order ot make a prediction. The consequence is that testing k-NN becomes slow and computationally expensive. This means that it is inappropriate for real-time application [5] but many researches reported that it has fairly high classification performance. Therefore, k-NN functions as good benchmarks for evaluating other classification models.

4.2 Naive Bayes Classifier

Naive Bayes classifier is included in probabilistic models as it computes both prior and posterior probabilities of the classes when it is given an input data whose class needs to predict. It is also identified as a generative classifier which learns a model of the joint probability, $p(x, y)$ of the inputs x and the label y [21]. In making prediction, the Bayesian rules are used to calculate $p(y|x)$ in order to pick the most likely label y. To calculate a class's prior probability, the distribution of features and parameters need to be estimated. Depending on how these parameters are estimated, the Naive Bayes classifiers have several configurations such as Gaussian Naive Bayes (GNB), Bernoulli Naive Bayes, or Multinomial Naive Bayes. Since this study uses Gaussian distribution, the discussion will be focused to GNB.

As in other Bayes classifiers, the computation of posteriori probability in GNB uses Bayes formula as shown in Equation 3. It can be seen that in order to compute the posteriori probabilities, $P(c_i|x)$, it requires a prior probability $P(c_i)$ and the class-conditional probability distribution $p(x|c_i)$, where i=1, ..., N predefined classes. The class prior probability $p(c)$ is computed by dividing the number of instances of class c with the total number of instances in the dataset.

$$\begin{aligned} P(c_i|x) &= \frac{P(c_i)p(x|c_i)}{p(x)} \\ &= \frac{P(c_i)p(x|c_i)}{\sum_{j=1}^{M} P(c_j)p(x|c_j)} \end{aligned} \tag{3}$$

Then, the question is on how to compute the class-conditional probability. It is done by the help of Gaussian distribution as shown in equation 4.

$$P(x_i|c) = (2\pi\sigma^2_{x_i,c})^{\frac{1}{2}} * exp\left(-\frac{1}{2}\frac{(x_i - \mu_{x_i,c})^2}{\sigma^2_{x_i,c}}\right) \tag{4}$$

where μ refers to the mean, and σ signifies the training sample variance. The Equation 4 requires to compute the mean and variance first before computing the class-conditional probability.

4.3 Linear Discriminant Analysis

In spite of its name, Linear Discriminant Analysis (LDA) belongs to generative and parametric classifiers [5] instead of discriminative ones. Therefore, it considers the problem of predicting a class label y (prior probability) on the basis of a vector of features x_i, as in the Naive Bayes classifier. The generative approach assumes that $P(Y = 1) = P(Y = 0) = \frac{1}{2}$ and the conditional probability of X given Y is also computed by Gaussian distribution as in GNB [21]. Another assumption is that each feature or dimension has the same variance σ. The prediction of input data x is computed by applying Bayes' rule written as in Equation 5.

$$h_{Bayes}(x) = \underset{y \in \{0,1\}}{\operatorname{argmax}} P(Y = y)(P(X = x|Y = y)) \tag{5}$$

As Shwarzt and Ben-David wrote in [21], the log-likelihood is needed in order to predict $h_{Bayes}(x) = 1$. The Equation 6 shows how to compute this likelihood ratio.

$$llr = \frac{1}{2}(x - \mu_0)^T(x - \mu_0) - \frac{1}{2}(x - mu_1)^T(x - \mu_1) \tag{6}$$

Unlike the Gaussian NB, the main goal of LDA-based classifier model is to project the original dataset in d-dimensional onto a lower dimensional subspace, k-dimensional, where $k < d$. To achieve this goal, the computation of LDA is generally performed into the following steps:

(1) The first step is to compute the distance between the means of different classes from the dataset, which is called the *between-class matrix*
(2) The second step is to compute the within-class matrix which is a calculation of the distance between the mean and the samples of each class [23]. Both between-class and within-class matrices are addressed as a scatter matrix, W.
(3) The next step is to calculate the eigenvectors (V) along with their eigenvalues (λ) for matrix W.
(4) Then it proceeds with sorting the eigenvectors in descending order; the first k eigenvectors is used as a lower dimensional space (V_k) [23].
(5) The last step is to project all samples (X) onto the new subspace.

It is reported that LDA is a good model for dimensionality reduction and for the existence of some unlabeled dataset. However,

Tharwata et. al. [23] reported that LDA suffers from small sample problem. It easily fails to find the lower dimensional space if the dimensions are much higher than the number of samples in the data matrix.

4.4 Support Vector Machine-based Classifier

As a type of hyperplane classifier, Support Vector Machine (SVM) has been widely implemented for pattern recognition, included OCR, since the mid of 1990s and has contributed to countless state-of-the-art performances [5]. The main goal of SVM algorithm is to compute the maximal margin, or put it in a simple way, to draw a margin of some width to the nearest data points. The margin, which is a gap between the two closest class points, is closedly related to a hyperplane which separates data points having different class. The so-called support vectors are the data points which are closest to the margin of a hyperplane.

For binary classification, the decision function is computed by Equation 7.

$$f(x) = w^T x + b \tag{7}$$

where w is a weight vector, and b is a bias (also called as threshold). The distance between a point x and the hyperplane is defined as $w^T x + b = 0$. The margin obtains its largest distance to the positive class if $w^T x + b = +1$ and to the negative ones by $w^T x + b = -1$. The search for the maximum marginal hyperplane in SVM is done in 2 steps:

(1) Generate hyperplanes which segregates the classes in the best way in an iterative way. The goal is to minimize the error.
(2) Select the hyperplane with the maximum margin

In practice, the SVM algorithm is implemented with a kernel which takes a low-dimensional input space and expands it into a higher dimensional space. This is done by using a kernel trick. There are several kernels in SVM, such as linear kernel, Radial Basis Function (RBF), or polynomial kernel. The linear kernel is good for binary classification while RBF-kernel is projected to deal with the muti-class classification. The linear kernel function is defined in Equation 8, while the RBF-kernel is computed with the Equation 9.

$$k(\vec{x}_i, \vec{x}_j) = \varphi(\vec{x}_i).\varphi(\vec{x}_j) \tag{8}$$

where $\varphi(\vec{x}_i)$ is the feature vector in the expanded feature space and may have infinite dimensionality [5].

$$k(\vec{x}_i, \vec{x}_j) = exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2) \tag{9}$$

where γ is a parameter ranging from 0-1. The higher value of γ will fit the data training, however it may cause an overfitting due to the increase of the number of support vectors [22].

5 THE PROPOSED METHOD

The highlight of this research is to build some classifier models and to experiment them on the data to get which model is the most suitable to our data. However, the recognition process could not stand alone but it should be done along other processes in OCR pipeline. Therefore, the flow of the processes in this research is described in Figure 4.

From this Figure, it can be seen that the system receives an input in the form of a scanned or photographed page of a book, text, or manuscript, or simply a text captured by the smartphone. This text image is then preprocessed. The segmentation process is performed on the level of lines and charaters. After having individual character per page, the process proceeds with extracting the features of each character. This process was applied to both training and testing process. The feature vectors of characters are then saved in a database as training data, while the feature vectors from the tested text image is used for doing a match in a classification way. After being trained, the classifiers would be validated and evaluated.

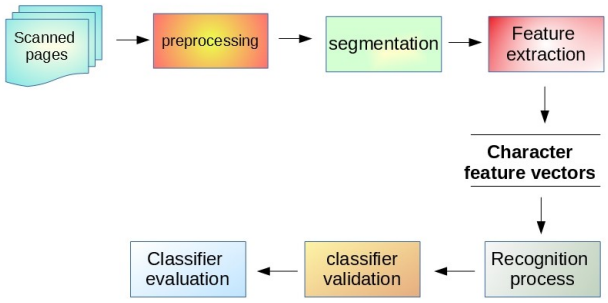


Figure 4: The architecture of the proposed system

5.1 Preprocessing Phase

Assuming that the input image has no geometric distortion and in relatively good condition, the preprocessing covers two tasks, i. e. the image gray-scaling and the background removal. The conversion of the input image into a grayscale one was implemented by using the function provided by opencv library in Python. In cv2, the function to load images in grayscale mode is callable through *cv2.IMREAD_GRAYSCALE*.

The background removal is a process which isolates the pixels forming the foreground, which is mainly text, from the pixels forming the rest of the image elements or the background. The foreground pixels are commonly marked with the integer 1, while the background pixels will be marked as 0. To remove the background, we simply applied the thresholding function provided by cv2 library in Python which is callable through *cv2.threshold*() syntax. We used the threshold $t = 160$ as it was suggested in [15]. However, we did an experimentation on 3 different threshold values, where $t = \{140, 150, 160\}$ to three different manuscripts printed on the end of 19^{th}, early 20^{th} centuries, and in 1970s. The nature of background color and print condition are totally different in these books. We decided to use $t = 160$ as it proves to give the best result to these three manuscripts, while $t = 140$ or $t = 150$ presents very poor and poor results to one of the manuscripts mentioned before.

5.2 Segmentation Process

In building our OCR system, we also applied the method of segmentation process suggested in [15] which concentrates more on solving the problem of line segmentation. The line segmentation for Javanese characters becomes a challenging task as it has upper and below elements of characters that are separated by the white spaces as shown in Figure 2. Applying the available segmentation methods

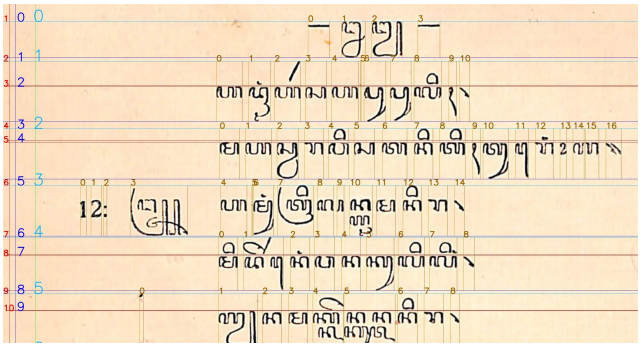


Figure 5: Three vertical lines on the left represent three passes on the line segmentation. Each line resulted from each pass is indexed and their difference on cutting the lines is marked by the index number.

will result in 2 or 3 separate lines which should be segmented into 1 line only. Therefore, the proposed method decomposes the line segmentation process into three stages or passes and one pass for the charater segmentation. Such segmentation strategy is aimed to adjust the nature of inputted text images having different height of line spaces. Figure 5 displays the line segmentation in three passes.

Briefly, the first pass of this three-stage line segmentation method is directed to find the line candidates by applying the Projection Profile Cutting (PPC) purely. In Figure 5, the PPC outputs are described with red lines and index numbers in red printed on the most left of the text. Having the line candidates, the task of the second pass is to remove falsely identified lines. This was done by removing any line candidades having the height value less than 5 pixels. The height threshold of 5 pixels came as a result of observation and experimentation applied on the segmentation of three manuscripts mentioned earlier. The ouput of the second pass is marked by the blue lines and index (see the middle vertical lines on the left on Figure 5).

The third pass is aimed to refine, split, or merge the falsely-identified line candidates by using the statistical technique to identify outliers of the vertical histogram of the text image. The information obtained from the statistical analysis was then used to regulate the thresholds and to decide whether to merge, split, or leave the lines as they were. The output of the third pass is displayed by the cyan horisontal lines, green vertical line (the third vertical line on the left of the text), and the cyan index numbers. Figure 5 shows that after a while, the third pass segmented the text into 6 lines shown by cyan index number 5, the second pass got 10 lines (index 9), and the pure PPC segmented the text into 11 lines (index 10).

After attaining lines of characters, the vertical segmentation was applied to obtain individual characters by analyzing the horizontal histogram of each line. This was done by applying Projection Profile Cutting technique. Figure 5 shows that mostly the characters are correctly segmented. The exception falls on those whose white space delimiter overlapped as shown by line index 5 and a set of characters which are indexed to number 5 also.

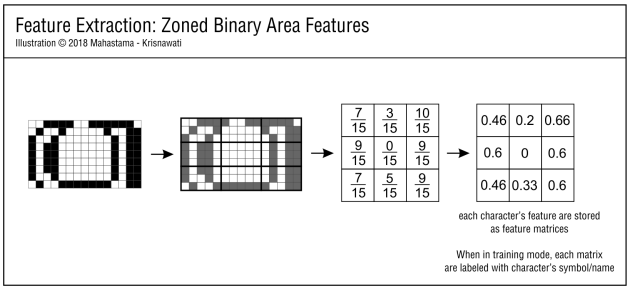


Figure 6: An illustration on how 9-area zoning works for extracting features of each character.

Table 1: An example of our feature extraction outputs in csv format.

f11	f12	f13	f21	f22	f33	f31	f32	f33	AS	class
.53	.25	.52	.57	.46	.37	.47	.35	.37	.31	na
.47	.58	.49	.13	.44	.41	0	.27	.39	.26	ra
.47	.07	.48	.51	.02	.48	.4	.33	.38	.26	pa
.54	.09	.23	.45	.17	.17	.61	.38	.33	.15	sa

5.3 Feature extraction Process

The feature of each character for both training and test data were extracted by applying a technique categorized as local feature extraction and a global one. The 9-area zoning was chosen to represent the local feature extraction methods. In this technique, the image is divided into 9 equal window sizes. For each window, the ratio of the foreground pixels to the whole pixels of each window was calculated and then saved into a $3X3$ matrix. The process of 9-area zoning is illustrated on Figure 6. From this technique, we got 9 features for each character.

As for the global feature extraction method, we applied the aspect ratio technique (AS) which is obtained by dividing the height of a character by its width. This technique results in a single feature value for each character. Combined with the 9 features attained from 9-area zoning, there are in total 10 features for each character to train and to match in the classification process. The extracted features were then saved into a csv file format which made the classification process easier. Table 1 exemplifies the csv file format of our feature extraction outputs wherein the first nine features are named according to the raw and the column in the matrix, AS stands for *aspect ratio* and the classes stand for the character classes.

5.4 The classification Process

A classification is a process of assigning a class for a given data. In practice, the class is a label given to a group of data points sharing the same characteristics which are then computed into the data point's feature vectors. In general, the task of a classifier model is described by the Equation 10, where y represents the output or the class of a given object x. Given a training set of labeled examples $(x_1, y_1), ...(x_n, y_n)$ and the object feature vector x, the prediction function f should be able to estimate to which class the object feature vector x belogs.

$$y = f(x) \tag{10}$$

In its implementation, the prediction function f takes form of a classifier. Each model of classifier attempts to predict the given object class using different approaches and techniques. We believe that there is no classifier inherently better than any other. For this reason, we built 4 different classifier models to experiment with our data. These classifiers are k-Nearest Neighbor, Gaussian Naive Bayes, Linear Discrimnant Analysis, and Support Vector Machine. The concept and algorithm of these classifiers have been explained in Section 4. We implemented the algorithms of these four classifiers using Scikit-Learn library in Python.

6 EXPERIMENT AND DISCUSSION

In order to do experiment and evaluate the prformance of these different classifier, we needed training as well as test data. Therefore, this section will present the strategy on how to get the data set, then it proceeds with the presentation of the experimental results and its evaluation.

6.1 Javanese Character Datasets

Since Latin alphabet was introduced by the colonial authority in 19^{th} century, the use of Javanese script in everyday life kept decreasing. Nowadays, it is no longer used but it can be found on the school lesson books and on the street signs in Yogyakarta and few cities in central Java. For this reason, our dataset was acquired by scanning 3 different books printed in different eras and which represent 3 different Javanese fonts. The first is the classical book *Serat Mangkunegaran IV* which was printed at the end of 19^{th} century, then the first 3 pages of book Matthew from the New Testament printed on early 20^{th} century and *Moelang Matja* which is a reading lesson book in Javanese script for elementary school and which was printed in 1970s. We selected randomly 72 pages out of these three books with one criteria only that these pages were free from scanning noises. However, they still contain light noises such as small blots or salt and pepper noise. Each page underwent the preprocessing and segmentation whose methods have been elucidated in the previous subsections. In average, one page comprises 250 character segments, the short page may consist ca. 180 segments and the longest one has ca. 320 segments. In total, we obtained 19.112 segments of characters excluding the undersegmentation or oversegmentation cases.

The data annotation was performed through two different methods: a crowdsourcing, annotation and evaluation workshops. Since not many Javanese could read and write Javanese script any more, the crowdsourcing was announced on the social media account of Wikimedia Indonesia (WMID). Thus, the annotator candidates are those who have interest on preserving Javanese script and language. The annotator selection process was conducted by administering 5 Javanese phrases with five different levels of difficulty. The task is to transliterate them in Latin alphabet and only those whose score is greater than or equal to 80 were accepted as annotators. From this process, we got 94 annotators.

Besides, we conducted annotation and evaluation workshop for the Javanese culture partisans living nearby Yogyakarta. The goal

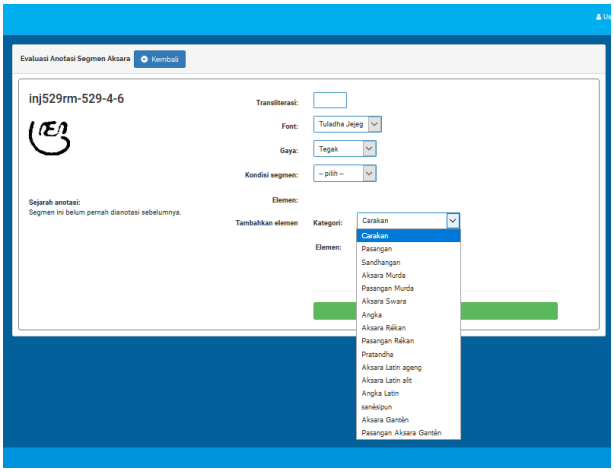


Figure 7: One of pages in annotation application Cakra developed for the prupose of dataset annotation and vaerification

was to to control the quality of annotation resulted from the crowdsourcing process. The evaluators were either volunteers or personnel of Javanese Wikipedia Community and could be considered as Javanese experts since most of them have degrees on Javanese literature or pedagogy. Through these techniques, we obtained 15.414 annotated characters and 633 classes whose correctness has been verified. The annotation and verification processes were enabled through a web-based annotation application *Cakra* that we developed for this specific purpose. One page of *Cakra* application is displayed in Figure 7.

6.2 Exprimental Results

In experimenting the four classifiers mentioned previously, we used 162 classes out of 633 available one. The reason is that more than half of these classes have less than or equals to 3 instances of annotated data. We considered that this is too small and decided to have minimally 10 instances of character samples in each class. The class in rank 162 is the cut-off for having minimally 10 annotated samples. These classes represent the frequently used characters which mostly are dominated by punctuations, *Carakan* and the most common compound characters. Naturally, these 162 classes excludes the archaic characters found only on the literary works such as *Serat Mangkunegaran*, and which in modern Javanese lesson books are no longer found.

For the experiments, we applied two phases of evaluation to our classifier models. The first is the model validation which evaluates our classifier models during the learning process. Based on the number of data, we opted 5-fold cross validation with a goal to get the most reliable model and which is able to accurately predict the classes in real use case. The second is the testing phase which assesses the real performance of the four models by the use of confusion matrix.

In validating our classifier models, we made use the class *KFold* provided by sklearn library in Python. We set $k = 5$ and splitted the dataset into a ratio of $80 : 20$, which did 5-fold cross

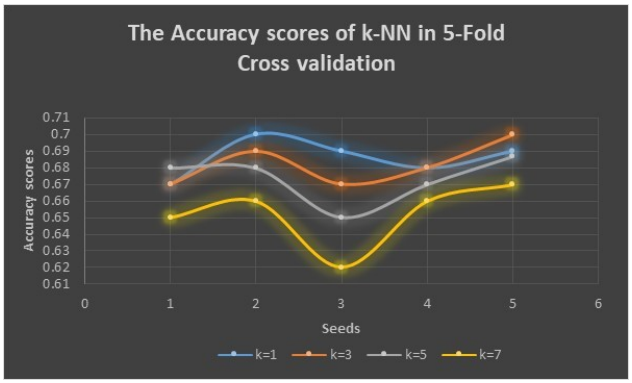


Figure 8: The accuracy scores of k-NN model in 5 seeds with $k = 1, 3, 5, 7$

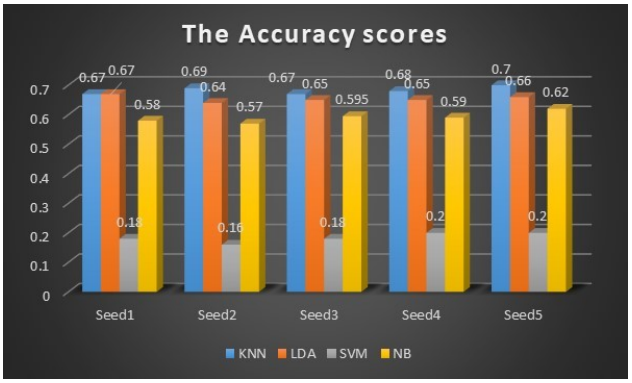


Figure 9: The accuracy scores of four experimented classifier models

validation by training each model 5 times on 80% of dataset and testing the rest, namely the 20%. The test data in each fold was guaranteed to be different. At first, we did some trial and error in determining the training-test data ratio and came to the best result on 80/20. In this case, we validated the models by computing their accuracy scores on each fold.

For the k-NN model, we tuned the leaf size parameter of the KDtree to 162 which corresponds to the number of classes. We decided to give the higher weight for closer neighbours which was realized by tuning the parameter weight to the *'distance'* value. The distance metric was set to 2 which refers to the Euclidean distance instead of using Manhattan distance metric. We experimented k-NN with $k = \{1, 3, 5, 7\}$. The accuracy scores of k-NN model is presented in Figure 8. Seed, in this chart, signifies the 20% data split to be the test data.

This experiment results showed that the $k = 1$ gained the highest average of the accuracy score by 0.69. However, the score difference is insignificant as the average of accuracy score remains on 0.6. It can be seen also that the k-parameter and the accuracy score are inversely proportional because the increase on k increases the possibillity of having bias and thus decreases the prediction accuracy. The chart in Figure 8 shows also that the accuracy scores of seed 5 are relatively higher than the other.

The accuracy scores of 4 experimented classifier models could be found in Figure 9. Tested on the limited number of data sample, k-NN proves to predict a bit more accurately than the other models. The most noticeable is the accuracy scores of SVM model. With the accuracy average of 0.18, SVM performance is quite far behind the other models. Our experiment found the contrast result to the SVM performance reported in some researches such as [6, 22], which stated that SVM showed great performance. In case that we applied a false kernel, we modified the kernel from using linear kernel into RBF kernel. However, the accuracy average of SVM using RBF kernel is much lower which achieves 0.02. To answer our curiosity to SVM's performance, we tested it with the Iris dataset provided on Github. The accuracy score of our SVM classifier tested on Iris dataset achieved 0.9 which is more than satisfying. We assumed that the underlying cause of poor performance of SVM is the insufficient number of training data. In this case, the number of features for each character sample is greater than the number of training data in each class. The ratio of the number of features to the number of training data in each class is proportional to 80% : 100%.

In implementing Linear Discriminant Analysis (LDA) model, we applied Singular Value Decomposition (SVD) for the solver parameter. The advantage of using SVD is that it does not compute the covariance matrix so that it saves the computation time. The prediction accuracy of LDA is competitive to k-NN in any seed (cf. Fig. 9) as the difference is less than 0.05. The Gaussian Naive Bayes (GNB) is on the third rank with the accuracy average on 0.59.

In testing phase, we evaluated the performance of the experimented models with the precision and sensitivity metrices. Unlike accuracy, the precision metric evaluates the prediction results on the basis of those predicted positive and hence precision is viewed as a finer metrics than accuracy in clasiffication case. In addition to precision, the proportion of actual positive cases that got predicted as positive (or true positive) was measured also using sensitivity metric. This metric is also well know as recall in the Information Retrieval concept. We applied micro average precision and sensitivity and compared them with the weighted ones. The Micro-Average Precision and Sensitivity are computed using the Equation 11 and Equation 12.

$$prec_{micro} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FP_i)} \tag{11}$$

$$sensi_{micro} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FN_i)} \tag{12}$$

where TP refers to True Positive prediction, FN stands for False Negative, and FP is for False Positive prediction.

The test data used in this phase were the same one used in the validation phase, which is approximately 20% of the dataset. Based on the confusion matrix, the precision and sensitivity of each tested data were able to compute. As presenting the whole matrix of evaluation results is impossible here, Table 2 displays a fraction of precision and sensitvity scores of character class from the evaluation outputs of LDA classifier. The character classes displayed in Table 2 were randomly chosen to represent different scores of precision and sensitivity.

Table 2: A fraction of precision and sensitivity scores on some character classes from evaluating the LDA classifier. The classes of characters are represented with the latin alphabets.

Classes	Precision	Sensitivity	Classes	Precision	Sensitivity
bi	1.00	0.50	dha	0.00	0.00
jar	1.00	1.00	mar	1.00	0.67
lê	0.00	0.00	ra	0.80	1.00
kir	0.67	0.50	kka	1.00	1.00
ma	0.00	0.00	mba	1.00	0.67
mê	0.33	0.50	ntu	0.67	1.00
pra	0.50	1.00	si	0.67	0.67
tri	1.00	1.00	wi	0.25	1.00

Table 3: The scores of micro-average precision, micro-average Sensitivity, and weighted-average precision and sensitivity of classifier models

	Micro-average		Weighted-averaged	
	Precision	Sensitivity	Precision	Sensitivity
KNN	0.69	0.69	0.74	0.69
LDA	0.66	0.66	0.71	0.66
SVM	0.18	0.18	0.18	0.18
GNB	0.62	0.62	0.67	0.62

The average Precision and Sensitivity were also weighted using the support values which resulted in the weighted average Precision and Sensitivity. Table 3 presents the measures on micro-average precision and sensitivity along with their weighted ones. This evaluation showed that the micro-average precision (MAP) scores of each model equal to their scores of micro-average sensitivity (MAS). Interestingly, the weighting technique has no influence at all on the weighted-average sensitivity (WAS) scores. However, it proves to be able to increase each model's precision score on the weighted-average precision (WAP). The exception falls to the SVM model whose MAP, MAS, WAS and WAP scores remain constant on 0.18. Experimented on our data, the SVM performance remains a mystery as we did not find a way to increase either its accuracy, MAP, or WAP. However, we assumed that the number of our annotated dataset is too small for training the SVM model.

As we scrutinized the confusion matrix, we found out that there are 6 characters in total which are unrecognizable by all classifier models and in every fold of 5-fold cross validation technique. Their precision rates are completely zero. Transliterated in Latin alphabets, this characters comprises *dha, ku, l, la, ma*, and the number *1* in Javanese. The failure to recognize number 1 in Javanese character is reasonable since its form and stroke are almost similar to the character *ga*, while there is no reasonable explanation why the other characters are always unrecognizable. There are also some characters whose precision and sensitivity scores are very low and their scores are no greater than 0.20 or 0 in most experiment folds. These characters are *2, lu, mu* and *wê*. We assumed that this low recognition rate for the number 2 in Javanese is that it has stroke which is almost similar to character *lê*.

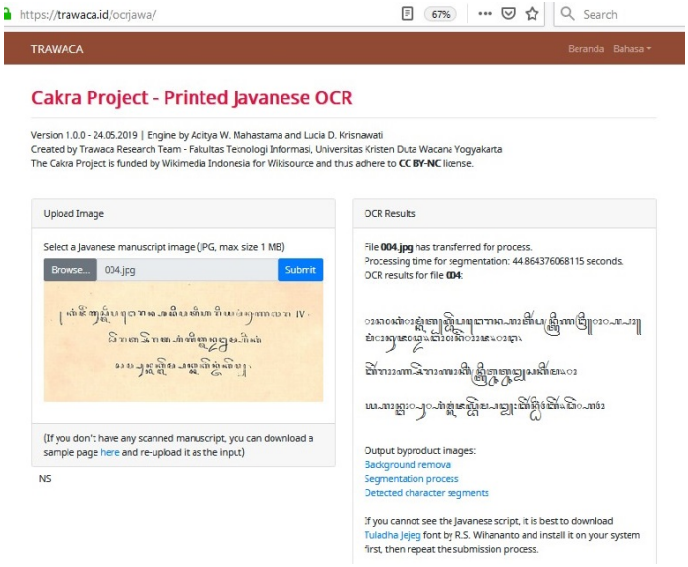


Figure 10: The capture of the on-off OCR system for Javanese characters

Though the accuracy, precision, and sensitivity scores of the experimented classifier models are still far from satisfying, the OCR system for Javanese character has been deployed and uploaded to the Trawaca website. It is accessible on the following link https://trawaca.id/ocrjawa. Processing the data online, it needs ca. 2-5 minutes to process the whole page of a Javanese text image containing ca. 250-300 characters. A sample page is provided also in case the reader has no scanned Javanese text. The byproducts of the processing stages such as background removal, segmentation and detected character segments could be seen by clicking the provided links. The capture of the on-off OCR system could be seen in Figure 10.

7 CONCLUSION

In this paper, we have presented the building process of four classifier models in a pipeline of an on-off printed OCR system for recognizing Javanese characters. These classifier models have been validated using 5-fold cross validation and assessed using the accuracy, precision and sensitivity metrices. Trained with characters taken from 3 books from different eras, the k-NN-based classifier outperforms any other classifier models. Its weighted-averaged precision score reaches 0.71 and its weighted-averaged sensitivity reaches 0.69.

Interestingly, the performance of SVM is very disappointing as its accuracy, precision, and sensitivity scores are averagely less than 0.2. In general, we assumed that the insufficient number of training data may cause the poor performance of SVM and prevents the other classifier models to attain the maximal performance on making a class prediction. This is proved by validating and testing our four classifiers to Iris dataset which resulted in weighted-averaged precision higher than 0.9. However, this research has contributed on the finding the total number of distinct Javanese characters which achieves 633 classes. Further, if it is compared to the researches on

the same field and objects (cf. the last 2 paragraphs of of section 2), this project has moved a step forward by using 162 character classes and deploying the system for real use instead of building a prototype trained with the basic Javanese characters only (*Nglegena*).

For future work, we need to improve the method of dataset annotation so that most classes have a well balanced number of character samples. The disadvantage of the previous method is that it took all characters from 72 pages to annotate and thus created a gap on the character frequency or the number of samples. The most commonly used characters have more annotated data, while more than half of the classes have less than or equals to 3 characters samples only. We projected that on the future annotation scheme, we would like to make a selection on the basis of lines containing compound and rare characters. For the classifier models, we plan to tune up more parameters and use some additional features to improve their performance, especially the Support Vector Machine.

8 ACKNOWLEDGMENTS

We would like to give our deepest gratitude to Wikimedia Indonesia (WMID) which has funded this research and partially funded the conference participation. We also give our thanks to Javanese Wikipedia community for their contribution on evaluating the character segments. We would also like to show our gratitude to Informatics Department, Duta Wacana Christian University which partially funded our participation in this conference. We thank to the head of Javanese Wikipedia Indonesia for his assistance in coordinating the workshops for evaluating our datasets. Last but least, we thank to everyone who was involved in the data annotation through the crowdsourcing process.

REFERENCES

[1] Bea Alex, Claire Grover, Ewan Klein, and Richard Tobin. 2012. Digitised Historical Text: Does it have to be mediOCRe?. In *Proceedings of KONVENS 2012 (LThist 2012 workshop)*. Vienna, 401–409.

[2] Muhammad R. Asif, Qi Chun, Sajid Hussain, Muhammed S. Fareeq, and Subhan Khan. 2017. Multinational vehicle license plate detection in complex backgrounds. *Journal of Visual Communication and Image Representation* 6, C (2017), 176–186. https://doi.org/10.1016/j.jvcir.2017.03.020

[3] Youssef Bassil and Mohammad Alwani. 2012. OCR Post-processing Error Correction Algorithm Using Google's Spelling Suggestion. *Journal of Emerging Trends in Computing and Information S* 3, 1 (Jan. 2012). http://arxiv.org/abs/1204.0191

[4] Wojciech Bieniecki, Szymon Grabowski, and Wojciech Rozenberg. 2007. Image Preprocessing for Improving OCR Accuracy. In *MEMSTEC'07*. Lviv-Polyana, 75–80.

[5] Mohamed Cheriet, Nawwaf Kharma, Cheng-Lin Liu, and Ching Y. Suen. 2007. *Character Recognition Systems: A Guide for Students and Practiotioners* (1st. ed.). Wiley, New Jersey.

[6] Chandra K. Dewa, Amanda L. Fadhilah, and Afiahayati. 2018. Convolutional Neural Networks for Handwritten Javanese Character Recognition. *Indonesian Journal of Computing and Cybernetics Systems* 12, 1 (Jan 2018), 83–94. https://doi.org/10.22146/ijccs.31144

[7] Maruf A. Dhali, Seng He, Mladen Popović, Eibert Tigschelaar, and Lambert Schomaker. 2017. A Digital Palaeographic Approach towards Writer Identification in the Dead Sea Scrolls. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, Vol. 1. Porto, Portugal, 693–702. https://doi.org/10.5220/0006249706930702

[8] Rudy Adipranata Gregorius S. Budhi. 2015. Handwritten Javanese Characters Recognition Using Several Artificial Neural Network Methods. *Journal of ICT Research and Applications* 8, 3 (2015), 195–212.

[9] Karez A. Hamad and Mehmet Kaya. 2016. A Detailed Analysis of Optical Character Recognition Technology. *International Journal of Applied Mathematics, Electronics and Computers* 4 (2016), 244–249.

[10] Mahdi Hamdani, Haikal El Abed, Monji Kherallah, and Adel M. Alimi. 2009. Combining Multiple HMMs Using On-line and Off-line Features for Off-line Arabic Handwriting Recognition. In *10th International Conference on Document Analysis and Recognition*. IEEE Computer Society, 201–205.

[11] Brian Karundeng, Kho I. Eng, and Anto S. Nugroho. 2009. An Evaluation of Feature Extraction Aalgorithm for Automatic Language Transcription System for Acient Handwriting Javanese Manuscripts. In *Proceeding of d^{rd} International Seminar on Industrial Engineering and Management*. Bali, G59–G65.

[12] Lucia D. Krisnawati and Aditya W. Mahastama. 2018. A Javanese Syllabifier Based on its Orthographic Forms. In *2018 International Conference on Asian Language Processing, IALP*. Bandung, Indonesia, 244–249. https://doi.org/10.1109/IALP.2018.8629173

[13] Changqing Liu. 2012. On Tangut Historical Documents Recognition. In *International Conference on Medical Physics and Biomedical Engineering (Pysics Procedia)*, Vol. 3. Elsevier, 1212–1216.

[14] Jean-Marc Ogier Made W. A. Kesiman, Jean-Christophe Burie. 2016. A New Scheme for Text Line and Character Segmentation from Gray Scale Images of Palm Leaf Manuscript. In *15th InternationalConference on Frontiers in Handwriting Recognition 2016*. Shenzhen,China, 32–330.

[15] Aditya W. Mahastama and Lucia D. Krisnawati. 2019. Improving Projection Profile for Segmenting Characters from Javanese Manuscripts. In *International Conference on Intermedia Arts and Creative Technology (CreativeArts)*. Yogyakarta.

[16] R. Manmatha and Toni M. Rath. 2003. *Indexing of Handwritten Historical Documents - Recent Progress*. Retrieved May 18, 2019 from https://pdfs.semanticscholar.org/47a3/de4595eb1d486b9283e415356bc0c322b462.pdf

[17] Marcin Namysl and Iuliu Konya. 2019. Efficient, Lexicon-Free OCR using Deep Learning. *CoRR* abs/1906.01969 (2019). http://arxiv.org/abs/1906.01969

[18] Nazaruddin and Sayed Muchallil. 2017. Comparison Online to Offline Handwritten Jawi Character Recognition Application. *Indian Journal of Science and Technology* 10, 12 (2017), 1–5. www.indjst.org

[19] Toni M. Rath and R. Manmatha. 2003. Features for Word Spotting in Historical Manuscripts. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR '03)*. IEEE Computer Society, 218–222.

[20] Christian Reul, Marco Dittrich, and Martin Gruner. 2017. Case Study of a highly automated Layout Analysis and OCR of an incunabulum: 'Der Heiligen Leben' (1488). In *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2017)*. ACM, Ǵottingen, Germany, 155–160. https://doi.org/10.1145/3078081.3078098

[21] Shai Shalev-Shwartz and Shai Ben-David. 2014. *On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes*. Cambridge University Press, New York.

[22] Olarik Surinta, Mahir F. Karaaba, Lambert R.B. Schomaker, and Marco A. Wiering. 2015. Recognition of Handwritten Characters Using Local Gradient Feature Descriptors. *Engineering ApplicationsofArtificial Intelligence* 45 (2015), 405–414.

[23] Alaa Tharwata, Tarek Gaber, Abdelhameed Ibrahimd, and Aboul E. Hassanien. 2017. Linear Discriminant Analysis: A Detailed Tutorial. *AI Communications* 30, 2 (2017), 169–190. https://doi.org/10.3233/AIC-170729

[24] Anastasia R. Widiarti and Phalita N Wastu. 2009. Javanese Character Recognition Using Hidden Markov Model. *International Journal of Computer, Electrical, Automation, Control and Information Engineering* 3, 9 (2009), 2201–2204.