Machine Learning and Microsimulation Techniques on the Prognosis of Dementia: A Systematic Literature Review

Blekinge Institute of Technology

Department of Computer Science and Engineering (DIDD)

Department of Health (HIHA)

Version 3.0

February 2016

CHANGE RECORD

Date	Version	Description	Author(s)
2015-08-11	1.0	First version of the protocol, to be sent to the reviewers.	Ana Luiza D. Moraes
2015-09-01	1.1	Modifications on the Quality Assessment Checklist.	Ana Luiza D. Moraes
2015-09-15	2.0	Modifications on the inclusion criteria, Quality Assessment Checklist, Schedule and addition of the Data Synthesis and Reporting section.	Ana Luiza D. Moraes
2015-10-30	2.1	Changes in the search string.	Shahryar Eivazzadeh
2016-02-02	3.0	Correcting inconsistencies.	Ana Luiza D. Moraes

SUMMARY

1. BACKGROUND

The worldwide aging phenomenon is a reality and the proof is that life expectancy has been increasing over the years and has already surpassed the age of 75 in 66 countries, according to the World Health Organization statistics [1]. As the proportion of elderly individuals continually grows, the main health care concern switched from infectious diseases to chronic conditions, as most of them are strongly related to aging [2].

In the set of elderly individuals, aged 65 years or more, it is not uncommon the prevalence of more than one chronic condition, a situation where the term comorbidity or multimorbidity is applied. Even though the prevalence of comorbidities varies across the studies, it can be implied that it affects more than half of the elderly population, and this has social and economic implications as it is associated to individuals with some kind of disability and poor quality of life as well as a high usage of the public healthcare [3]. To illustrate this scenario, in 1999, a study conducted in the United States showed that almost half of the public healthcare users that aged 65 or higher were in comorbid situations and these cases accounted 89% of the stipulated healthcare budget [4].

As most of the health research is still focused on specific diseases [2][3][4], most of the clinical practice guidelines used by practitioners follow the same trend and very few of them address scenarios of comorbidities, making the care, treatment and assistance of patients in comorbid situations a challenge [5].

One challenge concerning the aging population's health that is also related to disabilities and high healthcare usage is dementia. This condition refers to the decline in the mental ability, severe enough to interfere with the patient's functional abilities and independence, and to which there is no disease modifying treatment available [6]. It is estimated that in 2010 over 36 million people were living in this condition and the trend suggests that this number might double in the next 20 years [7]. To aggravate this scenario, studies have shown that patients who suffer from dementia tend to have on average from 2 to 8 additional chronic diseases, characterizing a comorbid situation, which actually accelerates their decline in daily functioning [6][7].

Considering the implications that dementia and comorbidities bring to the quality of life of the patients, and also the economic impact where the number of individuals depending on the public health care rises, it is fundamentally important the research in these topics, as for improving the quality of the treatment individuals receive from a clinical point of view, as for the public health care management.

Fortunately, as the availability of electronic health data expands in volume, variety and the speed in which users can interact with it, data analysis methods can be used to support the research in health economics and epidemiology [8]. Therefore, techniques such as machine learning and microsimulation can be used to support the creation of knowledge-based frameworks that could

provide models, patterns, predictors etc, contributing in clinical decision support for clinicians, patients or individuals concerning patient-care and population health [9].

2. Research Protocol

This systematic review of the literature aims to investigate how data analysis methods, in special machine learning and microsimulation techniques, are being employed in the studies concerning the prognosis of dementia and comorbidities.

2.1 Research Question

The main question this study aims to answer is: **"How are the machine learning and microsimulation techniques being employed by the researches on the prognosis of dementia and comorbidities?".** This main question can be decomposed in the four questions below:

- Q1: Which machine learning and microsimulation techniques are being used in the dementia and comorbidities research?
- Q2: What data characteristics (variables, determinants and indicators) are being considered when applying the machine learning or and microsimulation techniques (physiological, demographic/social, genetics, lifestyle etc)?
- Q3: What are the goals of the studies that employ machine learning or microsimulation techniques for prognosis of dementia and comorbidities?
- Q4: Do the studies focus on individuals or populations?

2.2 Search Process

As for the search process used in this systematic review of the literature, automated searches will be run on *Pubmed*, *Web of Science* and *Scopus* digital databases.

The search strings to be used in the automated searches were structured using the PICO approach [11], in which the main question is decomposed in four parts: population of interest intervention, comparison and outcome. For this systematic literature review, the "comparison" component was not used due to its nature being a characterization. The components used for the automated searches are characterized below, and the change record of the complete search string with its pre-execution results can be found on the Appendix 1.

- **Population:** Studies that address researches on dementia and comorbidities.
- **Intervention:** Machine learning or microsimulation techniques.
- **Outcome:** Prognostic estimates on dementia and comorbidities at individual or population level.

2.3 PRIMARY STUDY SELECTION

The primary study selection will count with four participants, two evaluating the papers retrieved by the automated searches, and two that will act in conflict resolution. The figure 1 presents the steps to be followed, which are adapted from the PRISMA Statement four-phase process.

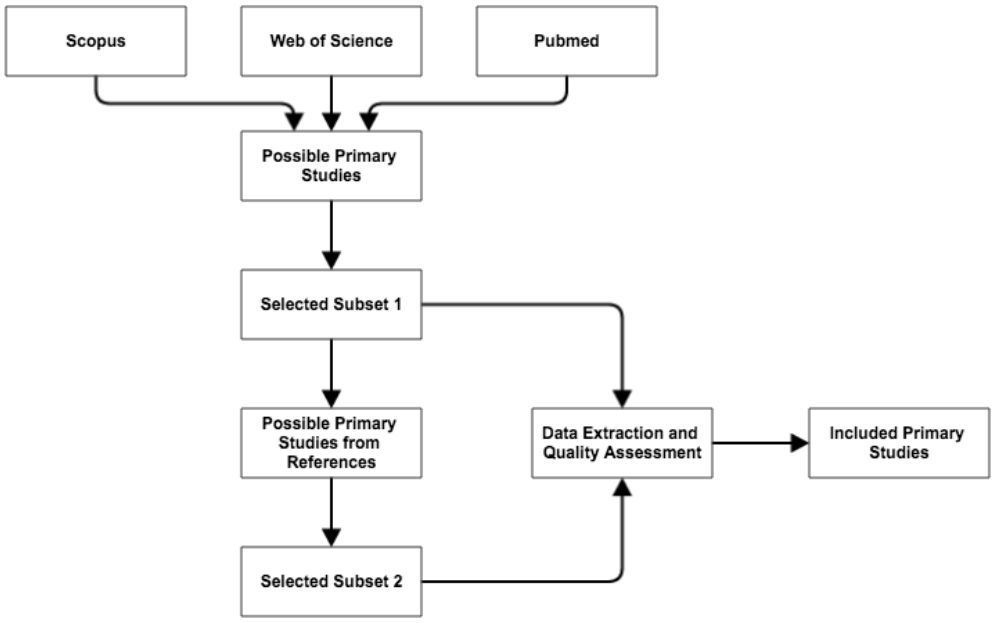

FIGURE 1. DIAGRAM FOR THE SYSTEMATIC LITERATURE REVIEW

Firstly, the results from the automated searches from the three selected sources (Scopus, Pubmed and Web of Science) will have their duplicates removed, and this set of possible primary studies will be assessed by the evaluation of the titles and abstracts according to the defined inclusion and exclusion criteria. The result from this evaluation is the Selected Subset 1. In the snowballing phase, the references list of the papers in the Selected Subset 1 will be evaluated, like in the previous phase and the resulting in the Selected Subset 2.

Each one of the papers contained in the Selected Subset 1 and 2 will be fully read and assessed for its quality using the quality assessment process specified in the section 2.4. The papers that successfully pass the established criteria will be part of the Included Primary Studies set and have its relevant data extracted.

Inclusion Criteria:

- Be a primary study in English; AND
- Address research on dementia and comorbidities; AND
- Address at least one machine learning or microsimulation technique; AND
- Address a prognosis related to dementia and comorbidities.

Exclusion Criteria:

- Be a secondary or tertiary study; OR
- Be written on another language other than English; OR
- Do not address a research on dementia and comorbidities; OR
- Do not address at least one machine learning or microsimulation technique; OR
- Do not address a prognosis related to dementia and comorbidities.

2.4 Study Quality Assessment

The papers selected to be fully read will have to pass through a quality assessment in which a quality checklist will be used. The checklist can be found in the table 1. Papers evaluated under 8 points will not be considered to be included in the set of primary studies. The conflicts found will be resolved in a meeting including all participants.

TABLE 1. PAPER QUALITY ASSESSMENT CHECKLIST ADAPTED FROM [12]

Design and Reporting of the Study	
Are the aims of the study clearly stated?	Yes=1 ; Partly=0.5 ; No=0
Does the study describe clearly the population being studied?	Yes=1 ; Partly=0.5 ; No=0
Was the sample size justified?	Yes=1 ; Partly=0.5 ; No=0
Is the sample representative of the population to which the results will generalize?	Yes=1 ; Partly=0.5 ; No=0
Were the limitations of the study reported either during the explanation of the study design or during the discussion of the study results?	Yes=1 ; Partly=0.5 ; No=0
Were the findings clearly reported?	Yes=1 ; Partly=0.5 ; No=0
Data Quality	
Are the measures used in the study clearly defined?	Yes=1 ; Partly=0.5 ; No=0
Are the measures used in the study valid?	Yes=1 ; Partly=0.5 ; No=0
Is the data collection method clearly described?	Yes=1 ; Partly=0.5 ; No=0
Technique(s) Employed	
Is/are the technique(s) being employed clearly described?	Yes=1 ; Partly=0.5 ; No=0
In case of more than one technique, was the statistical significance assessed?	Yes=1 ; Partly=0.5 ; No=0
Was a sensitivity analysis carried out to assess if the results were due to certain inputs?	Yes=1 ; Partly=0.5 ; No=0

2.5 Data Extraction

The papers that passed successfully through the quality assessment will have their relevant information extracted using the data extraction form that can be found in the table 2. All the participants will take part in this activity.

TABLE 2. DATA EXTRACTION FORM

ID				Extracted by	
Title				Journal/Source	
Authors				Year	
Publication Type				Country Origin	
Conditions Studied					
Database				Follow-up	
Dataset Categories					
Age Range			Gender		
Employed Data Analysis Techniques					
Physiological Variables					
Demographic/ Social Variables					
Genetics Variables					
Lifestyle Variables					
Other Variables					
Aim of the Study					
Focus					

2.6 Data Synthesis and Reporting

The data synthesis phase will compile the data extracted from the selected studies in order to answer the posed questions and to facilitate the following reporting process. This will be done tabulating the relevant extracted information for each question as shown in the tables 3,4,5 and 6.

- Q1: Which machine learning and microsimulation techniques are being used in the comorbidities and dementia research?

TABLE 3. SUMMARY OF EMPLOYED MACHINE LEARNING AND MICROSIMULATION TECHNIQUES

ID	Employed Techniques	Conditions Studied	Study Goals

- Q2: What data characteristics (variables, determinants and indicators) are being considered when applying the machine learning and microsimulation techniques (physiological, demographic/social, genetics, lifestyle etc)?

TABLE 4. SUMMARY OF DATA CHARACTERISTICS

ID	Variables					Conditions Studied
	Physiological	Demographic / Social	Genetics	Lifestyle	Other	

- Q3: What are the goals of the studies that employ machine learning or microsimulation techniques?

TABLE 5. SUMMARY OF STYDY GOALS

ID	Employed Technique	Conditions Studied	Study Goals	Follow-up

- Q4: Do the studies focus on individuals or populations?

TABLE 6. SUMMARY OF STYDY FOCUS

ID	Employed Technique	Conditions Studied	Focus

REFERENCES

[1] World Heath Organization. World Health Statistics 2015. http://www.who.int/gho/publications/world_health_statistics/2015/en/. Last accessed on 2015-06-19.

[2] Marengoni, Alessandra, et al. "Aging with multimorbidity: a systematic review of the literature." *Ageing research reviews* 10.4 (2011): 430-439.

[3] Melis, René, et al. "Incidence and predictors of multimorbidity in the elderly: a population-based longitudinal study." (2014): e103120.

[4] Boyd, Cynthia M., et al. "Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance." *Jama* 294.6 (2005): 716-724.

[5] Mutasingwa, Donatus R., Hong Ge, and Ross EG Upshur. "How applicable are clinical practice guidelines to elderly patients with comorbidities?." *Canadian Family Physician* 57.7 (2011): e253-e262.

[6] Melis, Rene JF, et al. "The influence of multimorbidity on clinical progression of dementia in a population-based cohort." (2013): e84014.

[7] Poblador-Plou, Beatriz, et al. "Comorbidity of dementia: a cross-sectional study of primary care older patients." *BMC psychiatry* 14.1 (2014): 84.

[8] Crown, William H. "Potential application of machine learning in health outcomes research and some statistical cautions." *Value in Health* 18.2 (2015): 137-140.

[9] Fraccaro, Paolo, et al. "Adoption of Clinical Decision Support in Multimorbidity: A Systematic Review." *JMIR medical informatics* 3.1 (2015).

[10] PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses. http://www.prisma-statement.org/statement.htm. Last Accessed on 2015-06-26.

[11] Pai, Madhukar, et al. "Systematic reviews and meta-analyses: an illustrated, step-by-step guide." *The National medical journal of India* 17.2 (2003): 86-95.

[12] B. Kitchenham. *"Guidelines for performing systematic literature reviews in software engineering (version 2.3)"*. Software Engineering Group, School of Computer Science and Mathematics, Keele University and Department of Computer Science, University of Durham, July 2007.

APPENDIX 1 – CHANGE RECORD OF THE PICO SEARCH STRING

String 1

- **Modifications:** Addition of the terms suggested in the last meeting and added the 10-year constraint to the results.
- **Pre Execution Results:**

Scopus	Pubmed	Web of Science
8003	2557	2953

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR **"Qualitative Research")**
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR **"Mixed Dementia"** OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR **"Frontotemporal Dementia"** OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
("Machine Learning" OR "Decision Tree" OR "Decision Support System" OR " Clinical Support System" OR "Neural Network" OR "Support Vector Machines" OR "Clustering" OR "Cluster" OR "Bayesian Network" OR "Genetic Algorithm" OR "Association Rule" OR " Regression")
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR "model" OR "patterns" OR **"diagnosis" OR "diagnostic"**)
)

String 2

- **Modifications:** Isolation of the keywords “Machine Learning” and “Data Mining” from the technologies parenthesis.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
22 (oldest from 2002)	6 (oldest from 2012)	7 (oldest from 2012)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR “Qualitative Research”)
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR “Mixed Dementia” OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR “Frontotemporal Dementia” OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
(("Machine Learning" OR “Data Mining”) AND ("Decision Tree" OR “Decision Support System” OR “ Clinical Support System” OR "Neural Network" OR "Support Vector Machines" OR "Clustering" OR "Cluster" OR "Bayesian Network" OR "Genetic Algorithm" OR "Association Rule" OR " Regression")
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR **“prediction“** OR "model" OR "patterns" OR “diagnosis“ OR “diagnostic” OR **“Forecasting”**))
)

String 3

- **Modifications:** Addition of the microsimulation keywords.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
22 (oldest from 2002)	7 (oldest from 2012)	9 (oldest from 2012)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR "Qualitative Research")
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining") AND ("Decision Tree" OR "Decision Support System" OR " Clinical Support System" OR "Neural Network" OR "Support Vector Machines" OR "Clustering" OR "Cluster" OR "Bayesian Network" OR "Genetic Algorithm" OR "Association Rule" OR " Regression") **OR "microsimulation" OR "micro-simulation" OR "microanalytic simulation")**
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting"))
)

String 4

- **Modifications:** Isolation of the keywords “Decision Support System” and “Clinical Support System” of the technologies parenthesis.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
32 (oldest from 2002)	7 (oldest from 2012)	9 (oldest from 2012)

- **Scopus String**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR “Qualitative Research”)
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR “Mixed Dementia” OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR “Frontotemporal Dementia” OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR “Data Mining” **OR “Decision Support System” OR “ Clinical Support System”**) AND ("Decision Tree" OR "Neural Network" OR "Support Vector Machines" OR "Clustering" OR "Cluster" OR "Bayesian Network" OR "Genetic Algorithm" OR "Association Rule" OR " Regression") OR “microsimulation” OR “micro-simulation“ OR “microanalytic simulation“)
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR “prediction“ OR "model" OR "patterns" OR “diagnosis“ OR “diagnostic” OR “Forecasting”))
)

String 5

- **Modifications:** Substitution of the technology's keywords for the ones in the ACM classification under "Machine Learning Approaches". The keywords were modified to remove plurals and words like "modeling", "learning" etc that were messing the results (returning an empty set).
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
38 (oldest from 2002)	7 (oldest from 2012)	13 (oldest from 2012)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR "Qualitative Research")
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "("Artificial life") OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation")
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting"))
))

String 6

- **Modifications:** Addition of other terms JBU suggested in the meeting and another keyword for microsimulation.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
38 (oldest from 2002)	7 (oldest from 2011)	12 (oldest from 2008)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR "Qualitative Research")
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "**Neurosyphilis**" OR "**complex of Guam**" OR **"Subcortical leukoencephalopathy"** OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR **"agent-based modeling"**)))
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting" OR **"projection"**))

String 7

- **Modifications:** Removal of the Outcome part of the string (the last parenthesis that refers to diagnosis, prognosis etc)
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
42 (oldest from 2002)	7 (oldest from 2011)	13 (oldest from 2008)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR "Qualitative Research")
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling"))))

String 8

- **Modifications:** Addition of the terms “population” and “epidemiological”
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
78 (oldest from 2002)	25 (oldest from 2004)	45 (oldest from 1999)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR “Qualitative Research” OR **“population” OR “epidemiological”**)
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR “Mixed Dementia” OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR “Frontotemporal Dementia” OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR “Data Mining” OR “Decision Support System” OR “ Clinical Support System”)
AND
(”Classification” OR ”Regression” OR ”Kernel” OR ”Support vector machines” OR ”Gaussian process” OR ”Neural networks” OR ”Logical learning” OR ”relational learning” OR ”Inductive logic” OR ”Statistical relational” OR ”probabilistic graphical model” OR ”Maximum likelihood” OR ”Maximum entropy” OR ”Maximum a posteriori” OR ”Mixture model” OR ”Latent variable model” OR ”Bayesian network” OR ”linear model” OR ”Perceptron algorithm” OR ”Factorization” OR ”Factor analysis” OR ”Principal component analysis” OR ”Canonical correlation” OR ”Latent Dirichlet allocation” OR ”Rule learning” OR ”Instance-based” OR ”Markov” OR ”Stochastic game” OR ”Learning latent representation” OR ”Deep belief network” OR ”Bio-inspired approach” OR ”Artificial life” OR ”Evolvable hardware” OR ”Genetic algorithm” OR ”Genetic programming” OR ”Evolutionary robotic” OR ”Generative and developmental approaches”)
OR
(“microsimulation” OR “micro-simulation“ OR “microanalytic simulation“ OR "agent-based modeling"))))

String 9

- **Modifications:** As suggested by PAN, this string contains only the comorbidities/dementia and machine learning/microsimulation synonims.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
546 (10-year range)	172 (10-year range)	354 (10-year range)

- **Scopus String:**

TITLE-ABS-KEY (
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling"))))

String 10

- **Modifications:** As suggested by EME, including the outcome parenthesis back back .
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
454 (10-year range)	143 (10-year range)	284 (10-year range)

- **Scopus String:**

TITLE-ABS-KEY (
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling")))
AND
("prognosis" OR "prognostic estimate" OR "conjecture" OR "conjecturing" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting" OR "projection"))

String 11

- **Modifications:** JBU suggestions. Removal of some synonyms that didn't make sense and restored the PICO string.
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
70 (oldest from 2002)	22 (oldest from 2004)	41 (oldest from 1999)

- **Scopus String:**

TITLE-ABS-KEY (
("longitudinal study" OR "longitudinal survey" OR "Observational Study" OR "cross-sectional study" OR "cross-sectional analysis" OR "transversal study" OR "prevalence study" OR "cohort study" OR "population" OR "epidemiological")
AND
("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity")
AND
((("Machine Learning" OR "Data Mining" OR "Decision Support System" OR " Clinical Support System")
AND
("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches")
OR
("microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling")))
AND
("prognosis" OR "prognostic estimate" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting" OR "projection")

String 12

- **Modifications:** The types of research was eliminated from the string, in sake of making sure all relevant results are included
- **Pre-Execution Results:**

Scopus	Pubmed	Web of Science
504	150	293

- **Strings:**

>>>>>>>>>>
Scopus:
>>>>>>>>>>

TITLE-ABS-KEY (("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity") AND ("Machine Learning" OR "Data Mining" OR "Decision Support System" OR "Clinical Support System") AND ("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches" OR "microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling") AND ("prognosis" OR "prognostic estimate" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting" OR "projection"))

>>>>>>>>>>
PubMed:
>>>>>>>>>>

("Dimentia" OR "Dementia" OR "Alzheimer" OR "Mixed Dementia" OR "Vascular Dementia" OR "Lewy Bodies" OR "Parkinson" OR "Creutzfeldt-Jakob" OR "Normal pressure hydrocephalus" OR "Huntington disease" OR "Wernicke-Korsakoff

Syndrome" OR "Frontotemporal Dementia" OR "Neurosyphilis" OR "complex of Guam" OR "Subcortical leukoencephalopathy" OR "Comorbidities" OR "Comorbidity" OR "Co-morbidity" OR "multimorbidity" OR "multimorbidities" OR "multi-morbidity") AND ("Machine Learning" OR "Data Mining" OR "Decision Support System" OR "Clinical Support System") AND ("Classification" OR "Regression" OR "Kernel" OR "Support vector machines" OR "Gaussian process" OR "Neural networks" OR "Logical learning" OR "relational learning" OR "Inductive logic" OR "Statistical relational" OR "probabilistic graphical model" OR "Maximum likelihood" OR "Maximum entropy" OR "Maximum a posteriori" OR "Mixture model" OR "Latent variable model" OR "Bayesian network" OR "linear model" OR "Perceptron algorithm" OR "Factorization" OR "Factor analysis" OR "Principal component analysis" OR "Canonical correlation" OR "Latent Dirichlet allocation" OR "Rule learning" OR "Instance-based" OR "Markov" OR "Stochastic game" OR "Learning latent representation" OR "Deep belief network" OR "Bio-inspired approach" OR "Artificial life" OR "Evolvable hardware" OR "Genetic algorithm" OR "Genetic programming" OR "Evolutionary robotic" OR "Generative and developmental approaches" OR "microsimulation" OR "micro-simulation" OR "microanalytic simulation" OR "agent-based modeling") AND ("prognosis" OR "prognostic estimate" OR "predictor" OR "prediction" OR "model" OR "patterns" OR "diagnosis" OR "diagnostic" OR "Forecasting" OR "projection")

>>>>>>>>>>
Web of Science:
>>>>>>>>>>

(TS="Dimentia" OR TS="Dementia" OR TS="Alzheimer" OR TS="Mixed Dementia" OR TS="Vascular Dementia" OR TS="Lewy Bodies" OR TS="Parkinson" OR TS="Creutzfeldt-Jakob" OR TS="Normal pressure hydrocephalus" OR TS="Huntington disease" OR TS="Wernicke-Korsakoff Syndrome" OR TS="Frontotemporal Dementia" OR TS="Neurosyphilis" OR TS="complex of Guam" OR TS="Subcortical leukoencephalopathy" OR TS="Comorbidities" OR TS="Comorbidity" OR TS="Co-morbidity" OR TS="multimorbidity" OR TS="multimorbidities" OR TS="multi-morbidity") AND (TS="Machine Learning" OR TS="Data Mining" OR TS="Decision Support System" OR TS="Clinical Support System") AND (TS="Classification" OR TS="Regression" OR TS="Kernel" OR TS="Support vector machines" OR TS="Gaussian process" OR TS="Neural networks" OR TS="Logical learning" OR TS="relational learning" OR TS="Inductive logic" OR TS="Statistical relational" OR TS="probabilistic graphical model" OR TS="Maximum likelihood" OR TS="Maximum entropy" OR TS="Maximum a posteriori" OR TS="Mixture model" OR TS="Latent variable model" OR TS="Bayesian network" OR TS="linear model" OR TS="Perceptron algorithm" OR TS="Factorization" OR TS="Factor analysis" OR TS="Principal component analysis" OR TS="Canonical correlation" OR TS="Latent Dirichlet allocation" OR TS="Rule learning" OR TS="Instance-based" OR TS="Markov" OR TS="Stochastic game" OR TS="Learning latent representation" OR TS="Deep belief network" OR TS="Bio-inspired approach" OR TS="Artificial life" OR TS="Evolvable hardware" OR TS="Genetic algorithm" OR TS="Genetic programming" OR TS="Evolutionary robotic" OR TS="Generative and

developmental approaches" OR TS="microsimulation" OR TS="micro-simulation" OR TS="microanalytic simulation" OR TS="agent-based modeling") AND (TS="prognosis" OR TS="prognostic estimate" OR TS="predictor" OR TS="prediction" OR TS="model" OR TS="patterns" OR TS="diagnosis" OR TS="diagnostic" OR TS="Forecasting" OR TS="projection")