



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique Et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère De L'Enseignement Supérieur Et De La
Recherche Scientifique



المدرسة الوطنية العليا للفلاحة الحراش – الجزائر
Ecole Nationale Supérieure Agronomique – El Harrach – Alger

Département : Génie rural
Spécialité : Sciences et techniques des agroéquipements

قسم الهندسة الريفية
تخصص علوم و تقنيات تجهيزات الفلاحة

Mémoire de fin d'étude En vue de l'obtention du Diplôme de Master

THEME

**Contribution to wheat grains weight estimation
using computer vision and machine learning**

Réaliser par: M.BOUKHELIF Abderazek
M.BISKER Mohamed Riadh

Soutenu le : 15/11/2021

Devant le jury composé de :

Encadreur : M. LAABASSI K.	MAA, ENSA
Président de jury : M.MEFTI M.	Pr, ENSA
Examineurs : M.ETSOURI S.	MCB, ENSA
Examineurs : M.BRAHIMI M.	MCB, Univ-BBA

Promotion 2016 – 2021

LIST OF CONTENTS

General introduction.....	14
I. Wheat origin and importance	17
I.1 The history of wheat.....	17
I.2 Bread and durum wheat taxonomic classification	18
I.3 Bread and durum wheat importance	19
II. Wheat production quantity and distribution in Algeria	20
II.1 Origin and diversity of wheat in Algeria	20
II.2 Production and distribution of wheat in Algeria.....	20
I.4 Registered Varieties of bread and durum wheat.....	23
III. Growth Stages of Wheat Plant	24
IV. The morphological description of a wheat plant parts	26
IV.1 The wheat spike description.....	26
IV.2 Structure of Wheat Grain.....	27
V. The Thousand Grain Weight	28
V.1 History and traditional ways of counting	28
V.2 Importance of the thousang grain weight.....	30
V.3 SOURCES OF GRAIN SIZE VARIATIONS.....	32
I. Introduction	34
II. What is Artificial Intelligence	34
II.1 Artificial Narrow Intelligence (ANI).....	34
II.2 Artificial General Intelligence (AGI)	34
II.3 Artificial Super Intelligence (ASI).....	34
III. Artificial narrow intelligence practical usage	35
III.1 Expert Systems.....	35
III.2 Machine Learning	35
III.3 Others ANI field application.....	36

IV. Machine Learning Regression Algorithms	38
IV.1 The Linear Regression:.....	38
IV.2 Decision Trees.....	39
IV.3 The Support Vector Machine.....	40
IV.4 The k- Nearest Neighbors	41
IV.5 The M5.Rules	42
IV.6 Multilayer Perceptron	43
V. Notion of distance	44
V.1 Euclidean Distance	44
V.2 Manhattan Distance	45
V.3 Minkowski Distance	45
V.4 Hamming Distance.....	45
VI. Machine learning process step by step.....	46
VI.1 A machine learning workflow	46
VI.1.1 Gathering Data.....	46
VI.1.2 Preprocessing and Processing Data	46
VI.1.3 Train Model.....	46
VI.1.4 Test Model	46
VI.1.5 Improve	46
VI.1.6 Prediction.....	46
VI.2 Machine learning problems and challenges	46
VI.2.1 Quality and Quantity of data	47
VI.2.2 Over and Underfitting	47
VI.2.3 Data labeling and annotating.....	48
VI.2.4 Bias	48
VII. Computer vision system methods related to object image analysis	48
VII.1 Image acquisition	49

VII.2	Preprocessing	49
VII.3	Image segmentation	50
VII.4	Feature Extraction	50
VII.5	Processing and decision-making or Classification	51
VIII.	Potential Machine Vision Systems in agriculture	51
VIII.1	Grading of Fruits and Vegetables	51
VIII.2	Sorting of Fruits and Vegetables.....	52
VIII.3	Related work to the Computer vision and Machine learning technologies in the agricultural field	52
I.	Introduction	57
II.	MATERIALS	59
II.1	Vegetal materials	59
II.2	Acquisition materials	60
II.3	Weighing material	61
II.4	The software	62
II.4.1	Computer vision and feature extraction	62
II.4.2	The Machine Learning Platform Weka	62
III.	Methods.....	63
III.1	The vegetal material collection.....	63
III.2	Image collection	64
III.3	Morphological features extraction	65
III.4	Data analysis using the platform Weka	66
IV.	Experimentation.....	68
IV.1	An overview of our workflow.....	68
IV.2	The feature selection step	70
IV.2.1	Correlation Based feature selection	70
IV.2.2	Learner Based feature selection	71
IV.2.3	Relief Based feature selection	73

IV.2.4	Features to keep and to remove.....	74
IV.3	Feature selection results summary.....	74
IV.4	The regression algorithms selection step	75
IV.4.1	Setup Weka	75
IV.4.2	Evaluating and comparing the results	76
IV.4.2.1	Correlation Coefficient (R)	77
IV.4.2.2	Mean absolute error (MAE)	77
IV.4.2.3	Root mean squared error (RMSE)	78
I.	Features selection results and discussion	80
I.1	CorrelationAttributeEval results.....	80
I.1.1	Full training set mode	81
I.1.2	Cross-validation mode	81
I.2	ReliefAttributeEval results	82
I.2.1	Full training set mode	82
I.2.2	Cross-validation mode	82
I.3	WrapperSubsetEval	83
I.3.1	Full training set mode	84
I.3.2	Cross-validation mode	84
I.4	Summary of the feature selection results	85
II.	The regression models selection process.....	85
	General Conclusion	89
	References	90
	Abstract.....	95

ملخص

الهدف من هذه الرسالة البحثية هو المساهمة في تقدير وزن الألف حبة (TGW) لأربعة أصناف من القمح وهي Vitron و Simeto للقمح الصلب، Arz و HD للقمح اللين من خلال الاعتماد على نموذج يتنبأ بوزن الحبة الواحدة (SGW). يستخدم النموذج الخصائص المورفولوجية المستخرجة بواسطة برنامج SmartGrain معتمداً على تقنية الرؤية الحاسوبية. إضافة على ذلك، تم إجراء هذا البحث للعثور على أفضل تصنيف انحدار متواصل والذي يعطي القيمة الأكثر واقعية لوزن حبة واحدة (SGW) مع أعلى معامل ارتباط R وأقل خطأ ممكن. تمر مهمة معالجة الصور بعدة خطوات: أولاً، تمت معالجة الحصول على الصور باستخدام ماسح ضوئي، ثم التجزئة واستخراج السمات (الصفات) بواسطة برنامج SmartGrain، ونموذج الانحدار التنبؤي بواسطة تقنية التعلم الآلي باستخدام منصة Weka. كانت الخصائص المورفولوجية (الميزات) هي مدخلات الخوارزميات LR و SVM و MLP و M5P.tree و M5Rules و Bagging هذه الأخيرة تعطي الأداء الأفضل في مهمة التنبؤ بدقة 93.72%. النتائج التي تم الحصول عليها من قبل المصنفات الأخرى كانت 93.42%، 93.36%، 93.14%، 92.93%، 92.93% على التوالي، وذلك للتنبؤ على مستوى الجنس (النوع) مع الخصائص المورفولوجية المختارة، باستخدام كلا الخيارين للاختبار لكونهما يعطيان نفس النتائج.

الكلمات المفتاحية: قمح؛ وزن ألف حبة؛ وزن حبة واحدة؛ الخصائص المورفولوجية؛ الرؤية الحاسوبية؛ معالجة الصورة؛ استخراج الميزات؛ نموذج الانحدار؛ التنبؤ؛ التقدير؛ التعلم الآلي؛ الذكاء الاصطناعي.

Résumé

Le but et l'objectif de cette thèse de recherche est de contribuer à l'estimation du poids mille grains (PMG) de quatre variétés de blé Vitron et Simeto pour le blé dur, Arz, et HD pour le blé tendre en s'appuyant sur un modèle prédisant le Poids de Grain Unique (PUG). Le modèle utilise les caractéristiques morphologiques extraites par un logiciel de vision par ordinateur « SmartGrain ». De plus, cette recherche a été menée pour trouver le meilleur classificateur de régression qui donne la valeur la plus réaliste du Poids de Grain Unique (PUG) avec le coefficient de corrélation R le plus élevé et le moins d'erreur possible. La tâche de traitement d'image passe par de nombreuses étapes : Tout d'abord, l'acquisition d'image a été abordée à l'aide d'un scanner, la segmentation et l'extraction d'attributs par le logiciel SmartGrain, le modèle de régression de prédiction par la technologie Machine Learning en utilisant la plateforme Weka. Les caractéristiques morphologiques étaient les entrées des algorithmes LR, SVM, MLP, M5P.tree, M5Rules.rules et Bagging. Le dernier classificateur est celui qui donne la meilleure prédiction avec une précision de 93,72%. Les résultats obtenus par les autres classificateurs étaient respectivement de 93,42 %, 93,36 %, 93,14 %, 92,93 %, 92,93 %, et ces résultats au niveau du Genre, avec les caractéristiques morphologiques choisies, en utilisant les deux options de test comme ils donnent les mêmes résultats.

Mots-clés : Blé ; Poids de Mille Grains ; Poids de Grain Unique ; Caractéristiques morphologiques ; Vision par ordinateur ; Traitement d'image ; Extraction de fonctionnalités ; Régression ; Prédiction ; Apprentissage automatique ; Intelligence artificielle.

Abstract

The aim and the objective of this research thesis is to contribute in the estimation of the thousand grain weight (TGW) of four varieties of wheat Vitron and Simeto for durum wheat, Arz, and HD for bread wheat by relying on a model that predict a Single Grain Weight (SGW). The model use the morphological characteristics extracted by a Computer Vision software "SmartGrain". Furthermore, this research was carried out to find the best regression classifier which gives the most realistic value of a Single Grain Weight (SGW) with the highest correlation coefficient R and the least possible error. The image processing task pass by numerous steps: First, the image acquisition was been tackled using a scanner, the segmentation and the attributes extraction by SmartGrain software, the prediction regression model by Machine Learning technology using Weka Platform. Morphological characteristics (features) were the inputs for the LR, SVM,

MLP, M5P.tree, M5Rules.rules algorithms and the Bagging one. The last classifiers, gives the best in the prediction task with an accuracy of 93.72%. The results obtained by the others classifiers were 93.42%, 93.36%, 93.14%, 92.93%, 92.93% respectively, and that for predicting at the Genus level with the choosen morphological characteristics, using both of the test option as they give the same results.

Keywords: Wheat; Thousand Grain Weight; Single Grain Weight; Morphological characteristics; Computer Vision; Image processing; Features extraction; Regression; Prediction; Machine Learning; Artificial Intelligence.