SOME NEW TYPES OF APPROXIMATIONS VIA MINIMAL STRUCTURE

M. M. El – Sharkasy and Fatma Altwme

1. Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.
   E-mail: sharkasy78@yahoo.com & mohamed.elsharkasi@science.tanta.edu.eg.
2. El- Zawia University – Libya.
   E-mail: fatmaaltwme@yahoo.com

Received 23/1/2018 Revised 28/2/2018

Abstract. In this paper, we use the notions of a minimal structure approximation space (short MSAS) and the notion of near open sets to introduce a new approximation of uncertain sets as a mathematical tool to modify the approximations. Relationships between these types are established via proof and counter examples. Also, some basic concepts of near approximations set are investigated and studied the relations between these different types of sets in MSAS. This set is a specific importance to help with the modifications of an approximation space via adding new concepts and facts. Finally, we use this concept to introduce the definitions of near lower approximation, near upper approximation, near boundary region, near rough and near exact sets and study some of the properties of this notion.

Keywords and phrases: Minimal structure, topological space, near open set, near closed set, Rough set and approximation space.

Mathematics subject classification: 54A05, 54C55, 54E05.

1. Introduction and Preliminaries

In 1963, Levine [1] introduced the notion of a semi-open (briefly s-open) set which is weaker than the notion of an open set in topological spaces. Since then several interesting generalized open sets have come into existence, which one of them is preopen set. The concept of a preopen set was introduced by Mashhour et al. [2]. In 2000, V. Popa and T. Noiri [3] introduced the notion of minimal structure. Also, they introduced the notion of MS-open set and MS-closed set and characterized these sets using MS-closure and MS-interior operator respectively. Rough set theory has been introduced in [4] as an extension of set theory. Many authors [4, 5, 6] introduced the relationship between rough sets and topology. Also, in [7, 8, 9, 10] using the concepts of topologies and the notion of minimal structure for processing data.

In this paper, some basic definitions and some properties of near open sets and near closed sets are studied in terms of PAS. Also, the relationship between these types is established via proof and counterexamples. The study and research about near open and closed sets have specific importance to help the modifications of the approximation space via adding new concepts and facts. We used minimal structure concepts to introduce definitions of near approximation and near boundary regions. We observe that the near approximations are mathematical tools to modify the approximations. Also, we introduce the near boundary regions as different areas of uncertainty. Finally, we used minimal structure concepts to introduce definitions of near rough and near exact sets.

Definition 1.1. [11] Let $(U, R)$ be a generalized approximation space where $U$ be a finite nonempty universe set and $R$ an arbitrary binary relation on $U$ and $N(x) = \{ y \in U : xRy \}$ is the right neighborhood of $x$ for all $x \in U$. Then the class $MS(U) = \{ \phi, U, N(x) \}$ is called a minimal structure on $(U, R)$. 
The members of the minimal structure $MS(U)$ are called $MS$-open sets and $(U, R, MS)$ be $MSAS$. The complement of an $MS$-open set is called $MS$-closed set, and let the class of all $MS$-closed sets will be denoted by $(MS(U))^c$.

The notions of lower approximation and upper approximation are the basic tools to define the sets. These operations have different forms and properties.

**Definition 1.2.** [11] Let $(U, R, MS)$ be $MSAS$ and $X \subseteq U$ then;
1. A minimal lower approximation of $X$ (in short $MS(X)$) is
   $$MS(X) = \cup \{G: G \in MS(X), \ G \subseteq X\}.$$
2. A minimal upper approximation of $X$ (in short $\overline{MS}(X)$) is
   $$\overline{MS}(X) = \cap \{F: F \in (MS(X))^c, \ X \subseteq F\}.$$

**Definition 1.3.** [12] Let $(U, R, MS)$ be $MSAS$. We say that $MS$ have a property of Maki if the union of any family of elements of $MS$ is in $MS$.

2. **Near open sets via MSAS**

In this section, we introduce some forms of near open sets and near closed sets via $MSAS$.

**Definition 2.1.** Let $(U, R, MS)$ be $(MSAS)$. We say that $A \subseteq U$ is;

1. A $MS$-regular open set if $A = MS(\overline{MS}(A))$. Also, we say that $A \subseteq U$ is an $MS$-regular-closed set if $A^c$ is an $MS$-regular open set.
2. A $MS$-semi-open set if there exist $X \in MS$ such that $X \subseteq A \subseteq \overline{MS}(X)$ equivalently, $A \subseteq \overline{MS}(MS(A))$. Also, we say that $A \subseteq U$ is an $MS$-semi-closed set if $A^c$ is an $MS$-semi-open set.
3. A $MS$-preopen set if $A \subseteq \overline{MS}(\overline{MS}(A))$. Also, we say that $A \subseteq U$ is an $MS$-preclosed set if $A^c$ is an $MS$-preopen set.

We denote the collection of all $MS$-regular open (resp. $MS$-regular closed, $MS$-semi-open, $MS$-semi-closed, $MS$-preopen and $MS$-preclosed) sets of $(U, R, MS)$ as $MS - RO(U)$ (resp. $MS - RC(U), MS - SO(U), MS - SC(U), MS - PO(U)$ and $MS - PC(U)$).

**Definition 2.2.** Let $(U, R, MS)$ be $MSAS$. We say that $A$ is nowhere dense if $MS(\overline{MS}(A)) = \emptyset$.

**Example 2.1.** Let $U = \{a, b, c\}$ and $R = \{(a, a), (b, b), (c, c), (a, b), (c, b), (a, c)\}$. Then $N(a) = \{U\}, N(b) = \{b\}, N(c) = \{b, c\}$. We obtain $MS(U) = \{\emptyset, U, \{b, c\}\}$ and $MS^c(U) = \{\emptyset, U \cup \{a, c\}, \{a\}\}$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$\overline{MS}(A)$</th>
<th>$MS(A)$</th>
<th>$\overline{MS}(MS(A))$</th>
<th>$MS(MS(A))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${a}$</td>
<td>${a}$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>${b}$</td>
<td>$U$</td>
<td>${b}$</td>
<td>$U$</td>
<td>$U$</td>
</tr>
<tr>
<td>${c}$</td>
<td>${a, c}$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>${a, b}$</td>
<td>$U$</td>
<td>${b}$</td>
<td>$U$</td>
<td>${b}$</td>
</tr>
<tr>
<td>${a, c}$</td>
<td>${a, c}$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>${b, c}$</td>
<td>$U$</td>
<td>${b, c}$</td>
<td>$U$</td>
<td>$U$</td>
</tr>
<tr>
<td>$U$</td>
<td>$U$</td>
<td>$U$</td>
<td>$U$</td>
<td>$U$</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
</tr>
</tbody>
</table>

Table 1

Then from Table 1 we have:
- $MS - RO(U) = \{\emptyset, U\}$,
- $MS - SO(U) = \{\emptyset, U, \{b\}, \{b, c\}\}$,
- $MS - PO(U) = \{\emptyset, U, \{b\}, \{a, b\}, \{b, c\}\}$,
- $MS - RC(U) = \{\emptyset, U\}$,
- $MS - SC(U) = \{\emptyset, U, \{a, c\}, \{a\}\}$.
"MS - PC(U) = {∅, U, {a, c}, {c, a}}.  

Example 2.2. Let U = {a, b, c, d} and R = {{a, a}, (b, b), (b, c), (c, b), (c, d), (d, c), (d, d)}. Then MS(U) = {∅, U, {a}, {a, b}, {b, c, d}}. By similar way 

MS - RO(U) = {∅, U, {a}, {a, b}, {b, c, d}}, 

MS - SO(U) = {∅, U, {a}, {a, b}, {a, c}, {a, b, c}, {a, b, d}, {b, c, d}}, 

MS - PO(U) = {∅, U, {a}, {a, c}, {a, b, c}, {a, b, d}, {c, d}, {a, b, c}, {b, c, d}}, 

MS - RC(U) = {∅, U, {a}, {a, b, c}, {a, c, d}, {b, c, d}}, 

MS - SC(U) = {∅, U, {a}, {a, b}, {a, c}, {a, b, c}, {a, b, d}, {b, c, d}}, 

MS - PC(U) = {∅, U, {a}, {a, b, c}, {a, c, d}, {b, c, d}}.  

Example 2.3. Let U = {a, b, c, d} and R = {{a, a}, (b, a), (b, c), (c, d), (d, c), (d, d)}. Then MS(U) = {∅, U, {a, b, c}, {c, d}}. By similar way 

MS - RO(U) = {∅, U}, 

MS - SO(U) = {∅, U, {a, c}, {a, b, c}, {a, b, d}, {b, c, d}}, 

MS - PO(U) = {∅, U, {a, c}, {a, b, c}, {a, b, d}, {b, c, d}, {a, c, d}, {a, b, c}, {b, c, d}}, 

MS - RC(U) = {∅, U}, 

MS - SC(U) = {∅, U, {a}, {b}, {d}, {a, b}}, 

MS - PC(U) = {∅, U, {a}, {b, d}, {b, c, d}}.  

Remark 2.1. In a topological space, the class of regular open sets is contained in the class of open sets. However, this fact was investigated for minimal structures and but surprisingly it is not true for the MS - RO(U) and MS(U) as the following example shows.

Example 5.2.4. Let U = {a, b, c, d} and R = {{a, a}, (b, b), (b, c), (c, d), (d, d)}. Then 

MS(U) = 

\[
\begin{array}{|c|c|c|c|c|}
\hline
A & MS(A) & \overline{MS}(A) & MS(\overline{MS}(A)) & \overline{MS}(MS(A)) \\
\hline
\{a\} & \{a\} & \{a, c\} & \{a\} & \{a, c\} \\
\{b\} & \{b\} & \{b, c\} & \{b\} & \{b, c\} \\
\{c\} & \emptyset & \emptyset & \emptyset & \emptyset \\
\{d\} & \emptyset & \emptyset & \emptyset & \emptyset \\
\{a, b\} & \{a, b\} & \{a, b\} & \{a, b\} & \{a, b\} \\
\{a, c\} & \{a, c\} & \{a, c\} & \{a, c\} & \{a, c\} \\
\{a, d\} & \{a, d\} & \{a, d\} & \{a, d\} & \{a, d\} \\
\{b, c\} & \{b, c\} & \{b, c\} & \{b, c\} & \{b, c\} \\
\{b, d\} & \{b, d\} & \{b, d\} & \{b, d\} & \{b, d\} \\
\{c, d\} & \{c, d\} & \{c, d\} & \{c, d\} & \{c, d\} \\
\{a, b, c\} & \{a, b, c\} & \{a, b, c\} & \{a, b, c\} & \{a, b, c\} \\
\{a, b, d\} & \{a, b, d\} & \{U\} & \{U\} & \{U\} \\
\{a, c, d\} & \{a, c, d\} & \{a, c, d\} & \{a, c, d\} & \{a, c, d\} \\
\{b, c, d\} & \{b, c, d\} & \{b, c, d\} & \{b, c, d\} & \{b, c, d\} \\
\emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\
U & U & U & U & U \\
\hline
\end{array}
\]

Table 2
And from Table 2, we have MS - RO(U) = {∅, U, {a}, {b}, {d}, {a, b}, {a, d}, {b, d}}, 

MS - SO(U) = {∅, U, {a}, {b}, {d}, {a, b}, {a, c}, {a, d}, {b, c}, {b, d}, {c, d}, {a, b, c}, {a, b, d}, {a, c, d}, {a, c, d}, {b, c, d}, {b, c, d}, {b, c, d}, {b, c, d}, {b, c, d}, {b, c, d}, {b, c, d}}, 

MS - PO(U) = {∅, U, {a}, {b}, {d}, {a, b}, {a, d}, {b, d}, {a, b, d}}, 

i.e., MS - RO(U) \not\in MS(U).  

3. Near lower and near upper approximation of uncertain concepts  
Lower and upper operators played a significant role in the approximation process. In this section, we introduce new types of lower and upper based on minimal structure approximation space. Also, we use of the regular open sets, semi-open sets, preopen sets to constructing approximation for uncertain concept of approximation space.
Definition 3.1. Let \((U,R,MS)\) be MSAS and \(A \subseteq U\). Then:
1. \(MS\) — regular lower approximation of a subset \(A\) (briefly \(MS_r(A)\)) defined by \(MS_r(A) = \cup \{ G : G \subseteq A, G \text{ is a } MS \text{ — regular open sets} \} \).
2. \(MS\) — regular upper approximation of a subset \(A\) (briefly \(MS_s(A)\)) defined by \(MS_s(A) = \cap \{ F : A \subseteq F, F \text{ is a } MS \text{ — regular — closed set} \} \).
3. \(A\) is regular exact if \(MS_r(A) = MS_s(A)\), otherwise \(A\) is said to be regular rough.

Example 3.1. From Example 2.2 we obtain:
If \(A = \{b, c\}\), then \(MS(A) = \{c\}\), \(MS_r(A) = \{b, c, d\}\), \(MS_s(MS(A)) = \{c, d\}\), \(MS(MS_r(A)) = \{b, c, d\}\), \(MS_r(A) = \{c\}\), and \(MS_s(A) = \{b, c, d\}\).

Proposition 3.1. Let \((U,R,MS)\) be MSAS and \(A \subseteq U\). Then:
1. \(MS(A) \subseteq MS_r(A) \subseteq A \subseteq MS_s(A) \subseteq MS(A)\).
2. \(MS(MS_r(A)) = MS_r(MS(A)) = MS_r(A)\).
3. \(MS(MS_s(A)) = MS_s(MS(A)) = MS(A)\).

Proof.
1. Since \(MS(X) \subseteq X\) and \(MS(X) \subseteq MS(MS(X))\). Then \(MS(X) \subseteq MS_r(X) = X \cap MS(MS(X))\). Also, since \(X \subseteq X \cup MS(MS(X))\) and \(MS(MS(X)) \subseteq MS(X)\). Then \(X \subseteq MS_r(X) \subseteq MS(X)\). \(2 - MS(MS_r(X)) = MS(MS(X) \cap MS(MS(X))) = MS(MS(X)) \cap MS(MS_r(X)) = MS_s(MS(X))\). Since \(MS(X) \subseteq MS(MS(X))\), then \(MS(X) = MS(MS_r(X)) \subseteq MS(MS(MS_r(X)))\) hence \(MS(MS_r(X)) = MS(X)\).

On the other hand \(MS(MS(X)) = MS(X) \cap MS(MS(MS_r(X))) = MS(X)\).
3. Similar to the proof of 2.

Definition 3.2. Let \((U,R,MS)\) be MSAS and \(A \subseteq U\). Then:
1. \(A\) is \(MS\) — semi lower approximation of a subset \(A\) (briefly \(MS_r(A)\)) defined by \(MS_r(A) = \cup \{ G : G \subseteq A, G \text{ is a } MS \text{ — semi — open sets} \} \).
2. \(A\) is \(MS\) — semi-upper approximation of a subset \(A\) (briefly \(MS_s(A)\)) defined by \(MS_s(A) = \cap \{ F : A \subseteq F, F \text{ is a } MS \text{ — semi — closed sets} \} \).
3. \(A\) is semi-exact if \(MS_s(A) = MS_r(A)\), otherwise \(A\) is said to be semi-rough.

Example 3.1. From Example 2.3, if let \(A = \{b, c, d\}\), then \(MS(A) = \{c\}\), \(MS_r(A) = U\), \(MS_s(MS(A)) = U\), \(MS_s(MS_r(A)) = U\), \(MS_r(A) = (b, c, d)\), \(MS_s(A) = U\).

The following propositions illustrate the properties of semi lower and semi-upper approximation.

Proposition 3.2. Let \((U,R,MS)\) be MSAS and \(X \subseteq U\). Then:
1. \(MS(X) \subseteq MS_r(X) \subseteq X \subseteq MS_s(X) \subseteq MS(X)\).
2. \(MS(MS_r(X)) = MS_r(MS(X)) = MS_r(X)\).
3. \(MS(MS_s(X)) = MS_s(MS(X)) = MS(X)\).

Proof.
1. Since \(MS(X) \subseteq X\) and \(MS(X) \subseteq MS(MS(X))\), then \(MS(X) \subseteq MS_r(X) = X \cap MS(MS(X))\). Also since \(X \subseteq X \cup MS(MS(X))\) and \(MS(MS(X)) \subseteq MS(X)\), then \(X \subseteq MS_s(X) \subseteq MS(X)\).
2. $\text{MS}(\text{MS}_S(X)) = \text{MS}(\text{MS}(X) \cap \text{MS}(\text{MS}(X))) = \text{MS}(\text{MS}(X)) \cap \text{MS}(\text{MS}(\text{MS}(X)))$. Since $\text{MS}(X) \subseteq \text{MS}(\text{MS}(X))$, then $\text{MS}(X) = \text{MS}(\text{MS}_S(X)) \subseteq \text{MS}(\text{MS}(\text{MS}(X)))$. Hence $\text{MS}(\text{MS}_S(X)) = \text{MS}(X)$. On the other hence $\text{MS}_S(\text{MS}(X)) = \text{MS}(X) \cap \text{MS}(\text{MS}(\text{MS}(X))) = \text{MS}(X)$. 

3. Similar to the proof of 2.

**Proposition 3.3.** Let $(U, R, MS)$ be MSAS and $X, Y \subseteq U$. Then:

1. $\text{MS}_S(\emptyset) = \emptyset, \text{MS}_S(U) = U$.
2. $\text{MS}_S(X \cap Y) \subseteq \text{MS}_S(X) \cap \text{MS}_S(Y)$.
3. $X \subseteq Y \Rightarrow \text{MS}_S(X) \subseteq \text{MS}_S(Y)$.
4. $\text{MS}_S(X \cup Y) \supseteq \text{MS}_S(X) \cup \text{MS}_S(Y)$.

**Proof.**

1. $\text{MS}_S(\emptyset) = \emptyset \cap \text{MS}(\text{MS}(\emptyset)) = \emptyset, \text{MS}_S(U) = U$ (obvious).
2. $\text{MS}_S(X \cap Y) = (X \cap Y) \cap \text{MS}(\text{MS}(X \cap Y)) = (X \cap Y) \cap \text{MS}(\text{MS}(X)) \cap \text{MS}(\text{MS}(Y)) = \text{MS}_S(X \cap Y)$.
3. $X \subseteq Y \Rightarrow \text{MS}(X) \subseteq \text{MS}(Y) \Rightarrow \text{MS}_S(X) \subseteq \text{MS}_S(Y)$.
4. $\text{MS}_S(X \cup Y) = (X \cup Y) \cap \text{MS}(\text{MS}(X \cup Y)) = (X \cup Y) \cap \text{MS}(\text{MS}(X)) \cup \text{MS}(\text{MS}(Y)) = \text{MS}_S(X) \cup \text{MS}_S(Y)$.

Another proof for 4. Since $(X \cup Y) \supseteq X$ and $(X \cup Y) \supseteq Y$, then by (3) we have $\text{MS}_S(X \cup Y) \supseteq \text{MS}_S(X)$ and $\text{MS}_S(X \cup Y) \supseteq \text{MS}_S(Y)$. Therefore $\text{MS}_S(X \cup Y) \supseteq \text{MS}_S(X) \cup \text{MS}_S(Y)$.

**Proposition 3.4.** Let $(U, R, MS)$ be MSAS and $X, Y \subseteq U$. Then:

1. $\text{MS}_S(\emptyset) = \emptyset, \text{MS}_S(U) = U$.
2. $\text{MS}_S(X \cup Y) \supseteq \text{MS}_S(X) \cup \text{MS}_S(Y)$.
3. $X \subseteq Y \Rightarrow \text{MS}_S(X) \subseteq \text{MS}_S(Y)$.
4. $\text{MS}_S(X \cap Y) \subseteq \text{MS}_S(X \cap \text{MS}_S(Y)) \supseteq (X \cap Y) \cup \text{MS}(\text{MS}(X \cap Y))$.

**Proof.**

1. Directly from Definition 3.2.
2. $\text{MS}_S(X \cup Y) = (X \cup Y) \cap \text{MS}(\text{MS}(X \cup Y)) = (X \cup Y) \cap \text{MS}(\text{MS}(X)) \cup \text{MS}(\text{MS}(Y)) = \text{MS}_S(X) \cup \text{MS}_S(Y)$.
3. Obvious (similar proof (3) in the Proposition 3.3)
4. $\text{MS}_S(X \cap Y) = (X \cap Y) \cap \text{MS}(\text{MS}(X \cap Y)) = (X \cap Y) \cup \text{MS}(\text{MS}(X \cap Y))$

291
\[
\begin{align*}
= (X \cup MS(MS(X))) \cap (Y \cup MS(MS(Y))) \cap (X \cup MS(MS(Y))) \cap (Y \cup MS(MS(Y))) \\
\subseteq (X \cup MS(MS(X))) \cap (Y \cup MS(MS(Y))) = MS(X) \cap MS(Y).
\end{align*}
\]

**Remark 3.1.** It should be noted that the property (2) in Proposition 3.3 and the property (2) in Proposition 3.4 different from Pawlak approximation.

**Proposition 3.5.** Let \((U, R, MS)\) be MSAS and \(X \subseteq U\). Then:
1. \(MS_3(X^c) = (MS_3(X))^c\).
2. \(MS_3(X^c) = (MS_3(X))^c\).
3. \(MS_3(MS_3(X)) = MS_3(X)\).
4. \(MS_3(MS_3(X)) = MS_3(X)\).

**Proof.**
1. \(MS_3(X^c) = (X^c) \cap MS(MS(X^c)) = (X^c) \cap MS((MS_3(X))^c) = (X^c) \cap MS((MS_3(X))^c) = (MS_3(X))^c\).
2. Similar to the proof of 1.
3. \(MS_3(MS_3(X)) = (X \cap MS(MS(X))) \cap MS(MS(X)) \subseteq MS_3(MS_3(X)) \subseteq MS_3(MS_3(X))\).

**Remark 3.2.** The properties (3) and (4) in the above propositions coincide with Pawlak approximation. Also, we have in general \(MS_3(MS_3(X)) \neq MS_3(X)\) and \(MS_3(MS_3(X)) \neq MS_3(X)\) which different from Pawlak approximation as the following example.

**Example 3.2.** From Example 2.2, let \(A = \{b, d\}\). Then \(MS_3(A) = \emptyset\), \(MS_3(A) = \emptyset\), \(MS_3(MS_3(A)) = \emptyset\), \(MS_3(MS_3(A)) = \emptyset\), \(i.e., MS_3(MS_3(A)) = MS_3(A)\) and \(MS_3(MS_3(A)) = MS_3(A)\) and \(MS_3(MS_3(A)) = MS_3(A)\).

**Theorem 3.1.** The set \(A\) is semi-exact iff \(MS(MS(A)) \subseteq A \subseteq MS(MS(A))\).

**Proof.** \(A\) is semi-exact. \(A = MS_3(A)\), \(A = MS_3(A)\)\(MS_3(A) = A \cap MS(MS(A)), MS_3(A) = A \cup MS(MS(A))\). Then \(A \cap MS(MS(A)) = A \cup MS(MS(A))\) but \(A \cap MS(MS(A)) \subseteq A\), \(A \subseteq A \cup MS(MS(A))\). Then \(A \cap MS(MS(A)) \subseteq A \subseteq MS(MS(A))\). From \(MS(MS(A)) \subseteq A \subseteq MS(MS(A))\), since \(A \cap MS(MS(A)) \subseteq A \subseteq MS(MS(A))\). Then \(A\) is semi-open and semi-closed, then \(A\) is exact.
In general the union of $MS$-open sets and $MS$—semi-open sets are not achieved. Nevertheless, for certain $MS$, the class of $MS$-semi-open sets are achieved under union of sets, like it is demonstrated in the following lemma.

**Lemma 3.1.** Let $MS$ be an minimal structure with satisfy of property of Maki and if $A_i \in MS - SO(U)$ for each $i \in I$. Then $U_{i \in I} A_i \in MS - SO(U)$.

**Proof.** Suppose that $MS$ has the property of Maki, and $A_i \in MS - SO(U)$ for each $i \in I$ there exists $G_i \subseteq A_i \subseteq MS$ such that $G_i \subseteq A_i \subseteq MS$ consequence, $U_{i \in I} G_i \subseteq U_{i \in I} A_i \subseteq U_{i \in I} MS(G_i)$. Since $MS$ is a monotone operator, then $U_{i \in I} MS(G_i) \subseteq MS(\bigcup_{i \in I} G_i)$ and $U_{i \in I} G_i \subseteq MS$. Because $MS$ has the property Maki. In consequence, $U_{i \in I} G_i \subseteq MS$ and $U_{i \in I} G_i \subseteq U_{i \in I} A_i \subseteq MS(U_{i \in I} G_i)$, then $U_{i \in I} A_i \in SO(U, R, MS)$.

As a consequence of the definition of $MS$-semi upper, we have the following result.

**Theorem 3.2.** Let $(U, R, MS)$ be MSAS on $U$ and $A \subseteq U$ then:

1. If $x \in MS(A)$, then $G \cap A \neq \emptyset$ for every $MS$-semi-open set $G$ such that $x \in G$.

2. In the case that $MS$ satisfies the property of Maki. Then $A$ is an $MS$-semi-closed set $\iff A = MS(A)$.

**Proof.**

1. If $x \in MS(A)$, then $\{F: A \subsetneq F, F$ is an $MS$ – semi closed set $\}$ and $x \in G$, then $G \cap A \neq \emptyset$ for every $MS$-semi-open set $G$ such that $x \in G$.

If $G \cap A \neq \emptyset$ for every $MS$-semi-open set $G$ such that $x \in G$, let $x \in A$ iff $x \notin U - A$ and $x \notin G \forall G \in MS(A)$: $G \subseteq U - A$ iff $x \in U - A = F, F \in (MS(A))'$: $A \subseteq F$ iff $x \in MS(A) = \cap \{F: A \subsetneq F, F$ is an $MS$ – semi closed set $\}$.

2. If $A$ is an $MS$-semi-closed set then $MS(A) = \cap \{F: A = F, A$ is a$MS$ – semi closed set $\}$ then $A = MS(A)$.

If $A = MS(A)$ then $MS(A) = \cap \{F: A = F, F$ is an $MS$ – semi closed set $\}$.

Then $A$ is a$MS$ – semi closed set.

**Theorem 3.3.** Let $(U, R, MS)$ be MSAS on $U$, $A \subseteq U$ and if $MS$ satisfies the property of Maki. Then $MS(A) = MS(A)$.

**Proof.** Since $MS$ satisfies the property of Maki. Then $MS(A)$ is an $MS$-semi-closed set. By using Definition 2.2, we obtain that $MS(MS(A)) \subseteq MS(A)$. Therefore, some new types $MS(MS(A)) \subseteq MS(A)$ and follows that $A \cup MS(MS(A)) \subseteq MS(A)$. The opposite inclusion, we observe that $MS(MS(A)) = MS(MS(A)) \subseteq MS(A)$. Thus, $MS(MS(A)) \subseteq MS(A) \cup MS(MS(A)) = MS(A)$. Following that, $MS(MS(A)) \subseteq A \cup MS(MS(A))$. As consequence by Definition 2.1 $A \cup MS(MS(A))$ is an $MS$-semi-closed set and so $MS(A) \subseteq A \cup MS(MS(A))$.

The following example shows that, if the property of Maki is removed in the previous theorem the equality is not necessarily true.
Example 3.3. Let \( U = \{a, b, c, d\} \) and \( R = \{(a, a), (b, a), (b, b), (c, c), (d, b), (d, d)\} \). Then \( MS(U) = \{\emptyset, U, \{a\}, \{b, c\}, \{a, b\}, \{b, c, d\}\} \).

\( MS'(U) = \{\emptyset, U, \{a, c\}, \{a, b, d\}, \{a, b, c, d\}\} \).

If \( A = \{b, d\} \), then \( MS(A) = \emptyset, MS_\emptyset(A) = \{b, d\}, MS_\emptyset(A) = \{a, b\}, MS_\emptyset(A) = \{a, b, d\}, \{b, d\} \subset \{a, b, d\} \), i.e., \( MS_\emptyset(A) \subseteq A \cup MS(\overline{MS}(A)) \).

**Theorem 3.4.** Let \((U, R, MS)\) be MSAS and \( A \subseteq U \). Then:

\[ MS_\emptyset(A) = A \cap MS(\overline{MS}(A)) \]

**Proof.** Follows from Theorem 2.2 and Theorem 3.2.

**Definition 3.3.** Let \((U, R, MS)\) be MSAS and \( A \subseteq U \). Then:

1. \( A \) is \( MS \)-pre lower approximation of a subset \( A \) (briefly \( MS_p(A) \)) defined by \( MS_p(A) = \bigcup \{G : G \subseteq A, G \text{ is a } MS \text{- preopen sets}\} \).
2. \( A \) is \( MS \)-pre upper approximation of a subset \( A \) (briefly \( MS_\emptyset(A) \)) defined by \( MS_\emptyset(A) = \bigcap \{F : A \subseteq F, F \text{ is a } MS \text{- preclosed sets}\} \).
3. \( A \) is pre exact if \( MS_p(A) = MS_\emptyset(A) \), otherwise \( A \) is said to be prerough.

**Proposition 3.6.** Let \((U, R, MS)\) be MSAS and \( A \subseteq U \). Then:

1. \( MS_\emptyset(A) = A \cap MS \overline{MS}(A) \).
2. \( MS_p(A) = A \cup \overline{MS} \overline{MS}(A) \).

**Proof.** Similar to the proof of Theorem 3.2.

The following proposition illustrates the properties of pre lower and pre upper approximations.

**Proposition 3.7.** Let \((U, R, MS)\) be MSAS and \( X \subseteq U \). Then:

1. \( MS(X) \subseteq MS_p(X) \subseteq X \subseteq MS_p(X) \subseteq MS(X) \).
2. \( MS(\overline{MS_p(X)}) = MS_p(\overline{MS(X)}) = MS(X) \).
3. \( MS(\overline{MS_p(X)}) = MS_p(\overline{MS(X)}) = \overline{MS(X)} \).

**Proof.**

1. Since \( MS(X) \subseteq X \) and \( MS(X) \subseteq MS(\overline{MS(X)}) \), then \( MS(X) \subseteq MS_p(X) = X \cap MS(\overline{MS(X)}) \). Also since \( X \subseteq MS(\overline{MS(X)}) \) and \( MS(\overline{MS(X)}) \subseteq MS(X) \), then \( X \subseteq MS_p(X) \subseteq MS(X) \).
2. \( MS(\overline{MS_p(X)}) = MS(X) \cap MS(\overline{MS_p(X)}) = MS(X) \cap MS(\overline{MS_p(X)}) = MS(X) \cap MS(X) \).
3. \( MS(X) \cap MS(\overline{MS(X)}) = MS(X) \).

4. **Near regions of uncertain concepts.**

In this section we obtain some rules to find near boundary regions in different ways in generalized approximation spaces with general binary relations.

**Definition 4.1.** Let \((U, R, MS)\) be MSAS and \( A \subseteq U \), then near exterior (\(i\)-exterior) of \( A \) is denoted by \( Ext_i(A) \) and is defined by \( Ext_i(A) = X - \overline{MS_p(A)} \), where \( i \in \{r, s, p\} \).

**Example 4.1.** From Example 2.2 the near lower and near upper approximations of \( A \) can be investigated as in Table 3.
Table 3

<table>
<thead>
<tr>
<th>A</th>
<th>$MS_i(A)$</th>
<th>$MS_r(A)$</th>
<th>$MS_s(A)$</th>
<th>$MS_p(A)$</th>
<th>$MS_p(A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
</tr>
<tr>
<td>{c}</td>
<td>{c}</td>
<td>{c}</td>
<td>{c}</td>
<td>{c}</td>
<td>{c}</td>
</tr>
<tr>
<td>{d}</td>
<td>$∅$</td>
<td>$∅$</td>
<td>$∅$</td>
<td>$∅$</td>
<td>$∅$</td>
</tr>
<tr>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
</tr>
<tr>
<td>{a, d}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
<td>{a}</td>
</tr>
<tr>
<td>{a, b, d}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
<td>{a, b}</td>
</tr>
<tr>
<td>{a, c, d}</td>
<td>{a, c}</td>
<td>{a, c}</td>
<td>{a, c}</td>
<td>{a, c}</td>
<td>{a, c}</td>
</tr>
<tr>
<td>{b, c, d}</td>
<td>{b, c}</td>
<td>{b, c}</td>
<td>{b, c}</td>
<td>{b, c}</td>
<td>{b, c}</td>
</tr>
</tbody>
</table>

Definition 4.2. Let $(U, R, MS)$ be MSAS and $A \subseteq U$, then near boundary be( i-boundary) region of $A$ is denoted by $BN_i(A)$ and is defined by $BN_i(A) = MS_i(A) - MS_s(A)$, where $i \in \{r, s, p\}$.

Definition 4.3. Let $(U, R, MS)$ be MSAS and $A \subseteq U$, then near positive ( i-positive ) region of $A$ is denoted by $POS_i(A)$ and is defined by $POS_i(A) = MS_i(A)$, where $i \in \{r, s, p\}$.

Definition 4.4. Let $(U, R, MS)$ be MSAS and $A \subseteq U$, then near negative (i -negative) region of $A$ is denoted by $NEG_i(A)$ and is defined by $NEG_i(A) = X - MS_i(A)$, where $i \in \{r, s, p\}$.

Theorem 4.1. Let $(U, R, MS)$ be MSAS and $A \subseteq U$, then $BN_i(A) \subseteq BN(A)$, for all $i \in \{r, s, p\}$.

Proof. obvious.

Example 4.2. In this example there exist the near regular (semi, pre) boundary region of $A$ by using Example 2.2 and Table 4.

<table>
<thead>
<tr>
<th>A</th>
<th>$BN_r(A)$</th>
<th>$BN_s(A)$</th>
<th>$BN_p(A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a}</td>
<td>$∅$</td>
<td>$∅$</td>
<td>$∅$</td>
</tr>
<tr>
<td>{b}</td>
<td>{b, d}</td>
<td>{b}</td>
<td>{b}</td>
</tr>
<tr>
<td>{c}</td>
<td>{d}</td>
<td>$∅$</td>
<td>{d}</td>
</tr>
<tr>
<td>{b, c}</td>
<td>{b, d}</td>
<td>{b, d}</td>
<td>{b, d}</td>
</tr>
<tr>
<td>{a, b, d}</td>
<td>{d}</td>
<td>$∅$</td>
<td>{d}</td>
</tr>
<tr>
<td>{a, c, d}</td>
<td>{b, d}</td>
<td>{b}</td>
<td>{b}</td>
</tr>
<tr>
<td>{b, c, d}</td>
<td>$∅$</td>
<td>$∅$</td>
<td>$∅$</td>
</tr>
</tbody>
</table>

Here we introduce the near roughness and near exactness by applying the concepts of near open sets. Let $A$ be a subset. Let $MS(A), MS_r(A)$ and $BN(A)$ be lower and upper and boundary region respectively. $A$ is exact if $BN(A) = ∅$ otherwise $A$ is rough. We shall express near rough set properties in terms of topological concepts.

Let $MS_i(A), MS_s(A)$ and $BN_i(A)$ be near lower, near upper and near boundary points respectively, where $i \in \{r, s, p\}$. $A$ is near exact (briefly i-exact) set if $BN_i(A) = ∅$, otherwise $A$ is near rough (briefly i-rough). It is clear $A$ is i-exact iff $MS_i(A) = MS_r(A)$. In pawlak space a subset $A \subseteq U$ has two possibilities rough or exact. The following definition introduces new types of definability for a subset $A \subseteq U$ in MSAS$(U, R, MS)$.

Definition 4.5. Let $(U, R, MS)$ be MSAS and $A \subseteq U$. Then:
1- $A$ is totally i-definable (i-exact) set if $MS_i(A) = A = MS_r(A)$.
2- $A$ is internally i-definable set if $A = MS_s(A)$.
3- $A$ is externally i-definable set if $A = MS_r(A)$.
4- $A$ is i-indefinable set if $A \neq MS_i(A), A \neq MS_r(A)$, where $i \in \{r, s, p\}$.
Proposition 4.1. Let \((U,R, MS)\) be MSAS and \(A \subseteq U\). So, if \(A\) is MS-exact than \(A\) is regular MS-exact.

Proof. Obvious by Definitions 1.2 and 3.1.

Proposition 4.2. Let \((U, R, MS)\) be MSAS and \(A \subseteq U\). So, if \(A\) is MS-exact than \(A\) is semi MS-exact.

Proof. If \(A\) is MS-exact, i.e. \(MS(A) = \overline{MS(A)}\). Then:
\[
MS_0(A) = A \cap \overline{MS\left(\overline{MS(A)}\right)} = A \cap \overline{MS\left(\overline{MS(A)}\right)} = A \cap \overline{MS(A)} = A.
\]
\[
MS_0(A) = A \cup MS\left(\overline{MS(A)}\right) = A \cup MS\left(\overline{MS(A)}\right) = A \cup MS(A) = A.
\]
i.e., \(MS_0(A) = \overline{MS_0(A)}\).

The converse of the above proposition is not true in general as the following example illustrates.

Example 4.3. If \(A = \{a,b,d\}\) in Example 2.2, then
\[
\overline{MS}(A) = \{a,b\}, \overline{MS}(A) = \{a,b,d\},
\]
i.e., \(\overline{MS}_0(A) = \overline{MS}(A) \neq MS(A).

Proposition 4.3. Let \((U, R, MS)\) be MSAS and \(A \subseteq U\). So, if \(A\) is MS-exact than \(A\) is pre MS-exact.

Proof. If \(A\) is MS-exact, i.e. \(MS(A) = \overline{MS}(A)\), then
\[
\overline{MS}_p(A) = A \cap \overline{MS\left(\overline{MS(A)}\right)} = A \cap \overline{MS\left(\overline{MS(A)}\right)} = A \cap \overline{MS(A)} = A.
\]
\[
\overline{MS}_p(A) = A \cup MS\left(\overline{MS(A)}\right) = A \cup MS\left(\overline{MS(A)}\right) = A \cup MS(A) = A.
\]
i.e., \(\overline{MS}_p(A) = \overline{MS}_p(A)\).

The converse of the above propositions is not true in general as the following example illustrates.

Example 4.4. If \(A = \{b,d\}\) in Example 2.3, then
\[
\overline{MS}(A) = \emptyset, \overline{MS}(A) = U, \overline{MS}_p(A) = \overline{MS}_p(A) = \{b,d\}
\]
and \(\overline{MS}(A) \neq \overline{MS}(A)\) i.e., \(\overline{MS}_p(A) = \overline{MS}_p(A) \neq \overline{MS}(A) = \overline{MS}(A)\).

References


[7]. Arpad Szaz, Minimal Structures generalized topologies and ascending systems should not be studied without generalized uniformities. Faculty of sciences and Mathematics University of Nis. Filomat, Vol. 21 (1), (2007), 87-97.


AUTOMATIC PLANT RECOGNITION AND DISEASES IDENTIFICATION METHODS BASED ON IMAGE PROCESSING TECHNIQUES

Ashraf Darwish
Faculty of Science, Helwan University, Cairo, Egypt
ashraf.darwish.eg@ieee.org

Received 17/1/2018 Revised Accepted 24/3/2018

Abstract
Plant recognition and diseases identification have an impact on the sustainable development of many countries in the agricultural sector. The automatic plant recognition and diseases identification will assist the specialists and experts in agriculture to overcome many of plant diseases and problems. The automation of plant diseases identification and recognition approaches have received considerable interest in the last years because their effect on the growth of the economy of countries, which may depend mainly on agriculture and to reduce the economic losses in the sustainable agriculture industry in general. However, human cognition and sight are not sufficient to identify the region of interest in the images of plants, usually, stems and leaves. Nowadays, image-based methods are considered as a visual assisting of plant recognition and diseases identification with the aid of the recent advances in image processing area. In this paper, we describe and analyze the automated image-based methods and discuss the state-of-art of plant recognition and diseases identification that has been applied in the last years. Also, we explore the role of image processing methods and classifiers in plant diseases identification and recognition. Different types of datasets of plant diseases identification and recognition are introduced briefly with their existing problems. As an example, the preprocessing phase of this issue is implemented based on real infected tomato leaves. Also, shape feature, color feature, and texture feature have been reviewed. Moreover, we described the important classifiers that are used currently used in the classification process. Also, hybrid classifiers can integrate the results from multiple algorithms with the aim of improving classification accuracy. Therefore, some of the well-known hybrid classifiers for plant diseases identification and recognition have been presented. Some solutions of using image-based methods such as complex backgrounds of the region of interest, different plant diseases can produce similar symptoms, and the conditions of capturing images have been presented. Finally, some points of the future work are proposed.

Keywords: Plant identification, Plant diseases, image processing, classifiers, feature extraction, shape feature, color feature, texture feature.

Mathematics Subject Classification: 68, 93, 94

I. Introduction and Related Work
Plant industry is representing the core of the growth of many countries. Usually, each plant has specific characteristics such as habits, morphology, and economic value. There are many kinds of plants worldwide which had been recorded and named by the statistics as in [1]. To make this industry effective, we apply image processing techniques to the study on plant recognition and identification. The developments in image processing area lead to used new methods and techniques in plant recognition diseases identification. With the advancements in digital pattern recognition and image processing methods, it will provide new opportunities for plant recognition and diseases identification. In the last decades, researchers gained an interest in plant diseases identification. Therefore, in the last years, researchers became interested in plant diseases identification for their importance and impact on the future of agriculture in adopting image processing techniques. However, plant diseases identification represents a new challenge in machine learning. The challenge in this process is the similarity in plant species which are very similar to each other. Also, plant maturity changes can lead to the large variation within one species.

The plant image identification issue was addressed as a track in the CLEF international conference as in [2]. In the literature, some systems had been developed for plant identification problems. For example, plant images for
identification and classification tasks as in [3, 4]. The richness of plant color is an important factor in identifying the
diseases. Therefore it is useful to depend on flowers plants for the identification process. In [5] different models of
color are used for the segmentation of flowers. According to [6], there are a number of fully automatic plant
identification systems in ImageCLEF’2012 that showed high efficiency in plant diseases diagnosis, LSIS-DYNI-run3
(LSIS - (laboratory des Sciences de l'Information et des Systèmes (http://www.lsis.org/), DYNI (Université du Sud
Toulon-Var, BP 20132, 83957 La Garde, France)), multiple classifications of plant leaves based on Gabor transform
and LBP operator are proposed in [7], and IFSC/USP (IFSC (Instituto de Física de São Carlos
(https://www2.ifsc.usp.br/english/)) [8]. Figure 1 describes the plant diseases identification based on image processin
two class problem techniques since 2000 to 2016 based on statistics of the web of knowledge databases.

Fig.1. Distribution of plant diseases identification from 2000-2016 according to Web of Science with some related
keywords.

The remaining of this paper is as follows. Section II presents the framework of plant recognition and diseases
identification, and some segmentations approaches are introduced. Section III explores the feature extraction methods
base on color, shape, and texture features. The commonly used classifiers in the literature are presented in section IV.
Section V presents the existing problems, limitations, and challenges in plant recognition and diseases identification.
Finally, the conclusion and the future work of this paper are described in section VI.

II. Image Processing Methods for Plant Identification Diseases

1. Traditional Methods of Plant Diseases Identification

The plant detections methods can be classified into two methods indirect and direct methods. In traditional methods,
there are serological (enzyme-linked immunosorbert assay, flow cytometer and molecular (fluorescence in situ
hybridization and DNA arrays). On the other hand, indirect methods include plant metabolite profiling, biomarker-
based detection of disease, fluorescence, and hyper spectral imaging techniques, and infrared RFID, sensors
techniques. Although the availability of such methods is still needed for low-cost methods, fast, and sensitive methods
to identify the diseases of plants. There are many automatic systems based on image processing for plant diseases
identification which will play an important role shortly in the protection of plant industry.

2. Why Computer-based images

Plants are an essential resource for foods to human especially with the increasing number of population and climate
change worldwide. Therefore, there is an interest to find new solutions based on the advances in computational
methods and techniques of an image processing to identify the plant diseases. The demand for identification of plant
with the help of computer vision and image processing has increasingly used in the last years to develop economies
of countries and to reduce the labor cost. In this regard, researchers developed a new automated system to identify the
plant diseases with a high level of accuracy. In the botany science, leaves of plants are representing an interest of
researchers because they can assist in differentiating different species of plant to identify the diseases. The
classification process is playing a crucial role in identifying plant diseases. With the developments in digital cameras,
this will increase demand for providing datasets of plant images that can be processed by image processing techniques.
Also, plant modeling is a complementary process in plant recognition and diseases identification process. In the
literature, there different algorithms that have been developed to reproduce the branching structure of plants as described in [9]. The different phases of plant identification process are described in Figure 2.

3. Image-based plant segmentation techniques

Plant image segmentation is based on the method of image acquisition. For example, scanning of separated leaf, photographing a plant leaf, photographing of a separated leaf with the complexity of background. In the segmentation process, isolated leaves that are photographed or scanned on a simple background are similar. In image segmentation process of a plant, the image can be segmented into two groups: plant and background, where background segmentation is an important process. In the literature, there are several methods that had been developed for segmenting the images of the plant. The segmentation methods are threshold-based segmentation, color index-based segmentation, and learning-based segmentation methods.

Image segmentation refers to discriminating objects of interest from their background. Threshold techniques are widely applied in plant detection based on image segmentation. These techniques pose two problems; namely, soil background class and plant vegetation class. Usually, thresholding is applied to transform the original image to determine the class. Choosing the proper threshold plays an important role in the segmentation process. Therefore, if the threshold value is set to be high value, then some important plant pixels may be merged with other background pixels which lead to under-segmentation. While a low threshold that is set to be low may lead to over-segmentation. In the literature, researchers have applied different threshold technique to address these problems. For example, authors in [10] presented a new retrieval method of images to plant identifying by using the features of contour-based shape. Moreover, authors in [11, 12] applied shape based leaf image retrieval method and leaf image retrieval with shape features for image retrieval problem.

Segmentation of the regions of the plant from the background consists of stones and soil is considered as an essential part to be performed for the detection of the diseases. Moreover, the performance of the segmentation process can affect the quality of the identified objects from the separated image information. Analyzing the behavior of color indices can provide an approximate understanding of the segmentation performance of the color indices. Their actual qualities cannot be drawn directly from the analysis unless the regions of plant and background are both known for a given context. In the color indices based approach, there are well-known color indices given in the literature which are; the normalized green, the difference between normalized green and normalized red, the normalized difference index, the excessive green index, the modified excessive green index, and the hue. All these indices use color components defined in the RGB or its normalized counterpart RGB space. However, other color models have also been used and tested in the literature. For example, hue, saturation, and intensity (HSI) color triplets transformed from the RGB system were employed to segment tomato plants from weeds [13] and for in-field weed sensing [14].

3. 1. Threshold-based approach

This technique has been applied in plant identification based on image segmentation has two group problems which are soil background class and plant vegetation class. Therefore, to remove the noise in the background of plant images, we can apply the threshold-based technique. In [15], authors presented a new method for threshold selection from gray-level histograms automatically. This method had been improved in [16] by using two-dimension Otsu method. Threshold-based additive Boolean update function attached to each element for the signaling network of a plant has been presented in [17].

3. 2. Color index approaches

To discriminate plants from the background, color is used as one of the most common methods for this purpose. Therefore, color distribution depends on color intensity, standard deviation, and skewness. Moreover, there are proposed image descriptors in the literature are Color Coherent Vector (CCV) [18], image indexing using Auto Color
Correlogram (ACC) [19], image retrieval method based on border/interior pixel classification [20], color indexing using GCH (Global Color Histogram) [21].

3.3. Learning based-approach

In some situations where we cannot perform the color-based approach in overcast and sunny conditions. Several studies have applied different approaches which include supervised and unsupervised machine learning methods with the transformation of color features. The classification process is based on the classifiers of machine learning such as Support Vector Machine (SVM), decision tree (DT), Naïve Bayes (NB), Naïve Bayes Tree (NBT). A comparative analysis of the leaf image recognition has been presented in [22] by using SVM classifier with 95.47% accuracy results. An expert system of plant identification has been proposed in [23] to identify different species of plant based on their images of leaves. In this system, the ant colony optimization algorithm is used for feature selection. The efficiency of this system was tested, and the average accuracy was 95.53%. A comparison of the recent method for the recognition of different plant species is described in [24].

3.4. Plant images data sets

In the last decade, there are multiple plant data sets are available for an attempt toward an electronic field guide for plants of researchers and specialists such as Smithsonian dataset [25], Swedish dataset for the classification process of leaves [26], ICL leaf dataset [27], and others. However, there is a problem of image datasets available for the implementation in plant identification. So far, there are a few number of used datasets are accessible to the scientific community or very limited. Currently, there exist some databases are available, such as the database developed in [28] to enable the development of mobile disease diagnostics through machine learning, in which roughly 53,000 images of healthy and diseased plants are available at (https://www.plantvillage.org/). Also, most of the exists data sets that are used in implementation in most of all researchers are collected manually using modern Cameras and images are captured by researchers themselves such as in CLEF conference in 2011 71 species of plant in 2011 that are increased to be 126 species in 2012 The datasets were collected in different categories based on scanned images of a leaf, scan-like photographs of a leaf and unconstrained photographs of plant leaf. In some cases, the image database can be created based on microscopes. The images of tomato plant that are introduced in this paper are captured from a farm by authors themselves. The implementation using Matlab toolbox of the preprocessing phase on tomato leaves plant is presented in Figure 3. This phase has been implemented according to the following steps.

1. Images Capturing. In this phase, SVM has been employed in this phase. In the acquisition process, we carefully selected images to achieve the classification process. The used images in this phase are captured to the infected leaves of tomato with two diseases, Powdery mildew, and early blight, from different farms. This real data set includes 200 infected leaves images were 100 for each virus type has been extracted.

2. Images Preprocessing. The main aim of this phase is removing noise in the image and enhancing smoothness that can be caused by the capturing conditions and environment. Therefore, we must isolate and extract each leaf in all images.

3. Image Resizing. Captured images are with different sizes and then these images resized to 512x512 resolution to reduce the size.

4. Background removal: In the background of images, a shadow may disturb the feature extraction phase. Therefore, the background of each image should be removed by using subtraction technique with some morphological operations.

![Fig. 3. Preprocessing phase of tomato leaves plant](image-url)

To analyze the results of this phase, the quality of the images has been enhanced by removing the noise that caused by defects of camera flash or high lights to increase the efficiency of the classification and prediction process.
Therefore, every leaf has been isolated with extract in the single image. Then, the captured images were resized to 512 x 512 resolution to minimize the storage capacity and reduce the computational time in the post-processing. Finally, the background of each image has been removed by using the background subtraction technique with some morphological operations. Gaussian Mixture-based Background/ Foreground Segmentation algorithm was used to subtract the background and morphological techniques (dilation followed by erosion) to remove noise.

III. Feature extraction in plant diseases
Feature extraction is an important stage in pattern classification. Nowadays, the datasets of plant images contain a lot number of features which sometimes are irrelevant or with the high dimension which can affect the performance of accuracy and effect of the results. Therefore, feature selection has been proposed in the last few years to improve the results of the used algorithms and techniques in plant diseases identification. The main objective of the feature selection process is to identify the most important features in the datasets for the optimization of the problems. Identifying the important subsets of the features is considered as important features in reducing the dimension of data to be processed in the classification process. Each image in a given dataset has its features and can be discriminated than other images by some features such as color, brightness, edge, and texture, etc. Other features such as histogram some can be obtained after processing.

Feature extraction is the most important stage of the diseases identification process in plants and segmentation process. For example, a new proposed method of the segmentation of the weed leaf by using deformable templates has been proposed in [29]. While a leaf extraction from complicated background is presented in [30], and a hybrid algorithm for leaf recognition and characterization of images is presented in [31].

4.1. Shape Feature
Shape feature is considered as an important feature in plant images, and the extraction process of shape features usually depends on the segmentation of the image. The region of interest (ROI) and the boundary of an object can be segmented.

In this scenario, the shape feature is divided into two groups: features based on regions and the boundary based features. Recently, researchers are applying the integration of shape, color, and texture features to improve the segmentation process. For example, authors in [32] classified plant leaves based on their texture features using Gabor features and LBP. Authors in [33] applied an integrated system which depends on using texture and shape features, while discrete wavelet transform (DWT) is applied to texture feature. Recognition of plant leaves using support vector machine has been presented in [34]. A shape classification method has been proposed in [35]. A new method for plant identification is presented in [36] based on multiscale fractal dimensions to identify the problem of leaf identification in the plant. Also, another leaf classification method is proposed in [37] based on a supervised method by using label propagation.

4.2. Texture Feature
The texture of an image is represented by the gray scale, and color change of the image is composed of many close elements. Therefore, the image texture describes the features of sparse, smooth, and so on. In the literature, the texture of an image can be classified into three methods which are structure method, statistics method, and frequency method. Usually, the texture is often overshadowed by the shape as the dominant for leaf recognition for example, in plant identification disease. The best texture descriptors in the literature are the Gabor features and local binary patterns and fractal dimensions [38]. There are state-of-the-art texture description methods which are introduced morphological texture descriptors as in [39, 40].

4.3. Color Feature
Color features are considered as visual features which are reliable and stable. The color is not discriminative as texture or shape in recognition of leaves, where plants usually are located in uncontrolled environmental conditions such as the shade of green. Moreover, the colors of leaves depend on the environmental conditions and the season of plants. Also, color is an important factor and has its contribution to the identification of plant diseases. For instance, authors in [41] proposed color descriptors such as RGB histogram, and LSH histogram and color moments [42].

IV. Imaging Classifiers for Plant Diseases Identification
Classification is an important process in data mining. A classifier uses the input datasets to build classification models. Classifiers usually are used for predicting data sets with binary classes. Using robust and efficient features are representing an important element in the classification process.

1. Support Vector Machine (SVM) Classifier
SVM classifier is widely used in different cases as in linear separable and linear inseparable by using the strategy of nonlinear mapping algorithm. The SVM objective is based on dividing the training set into two classes. The main idea of SVM is to map the original data into a high-dimensional feature and construct a hyper plane as the discriminative surface between the negative and positive data. SVM is used to solve the regression and classification
problems and in bioinformatics problems in different areas such as identifying the plant promoters [43]. SVM classifier is used in the literature to evaluate the effectiveness of each feature in groups and individually, during the implementation process. For evaluation process, authors used the remaining samples from the training set. Automation of plant cell detection process is an important to obtain the development information of individual plant cells to provide tools of research to ease the search for special events, such as cell division. SVM is used to select the cells based on the cell descriptor constructed from the edge strength and shape [44]. Authors in [45] designed an automated system to classify the wheat grains, and the results showed a high level of accuracy. SVM has been applied to differentiate three categories of tea plant; green, black and Oolong teas [46]. Herbal plant images are used in an automatic identification system as in [47] and SVM classifier is used to classify the herbal plants of licorice and rhubarb.

SVM classifier was implemented for an automated system of leaf plant recognition in [48] with an average accuracy of 95%. In this system, the proposed plant leaf classification technique used the SVM classifier that based on shape descriptor, and the rate of recognition was up to 95.16%. Authors in this work presented a method of plant species recognition using SVM for Flavia dataset. In the literature, there are other related models of plant identification and recognition based on leaves images as proposed in [49-53].

2. Artificial Neural Network (ANN)

ANN was used to train data set samples (Flavia and ICL) for the automatic classification of plants based on their leaves of a large number of different fruit trees and showed high level of accuracy [54]. Convolutional neural network (CNN) is used in some applications using Android mobile application to classify the natural images of leaves [55]. ANN was used in [56] to study the method of leaf identification with k-Nearest Neighbor (KNN) classifier with accepted recognition rate. ANN is used in leaf recognition with Probabilistic Neural Network (PNN) classifier for different classes of plant leaves samples, and the results showed high recognition rate of 93.08% [57].

2. k-Nearest Neighbor Classifier (KNN)

KNN is data mining classification techniques which is based on the closest training examples in the feature space can be represented by its closest K neighbors. A new method for plant leaves classification based on two-dimensional shape feature are proposed in [58]. In KNN, the classification of the selected data set can be performed by taking into account distances between data values of each class. In this context, the plant leaves classification, KNN classifier has been used to classify the plant species using a set of reduced leaf shape feature vectors. Leaf features are employed and used in different plants classification based on KNN classifier. Authors in [59, 60] employed KNN to discuss the leaf plant classification and the experimental results showed high recognition rate. A new proposed approach for leaf classification using local features of plant species from low quality pictures using mobile devices is presented in [61]. With the consideration of both local features and global features, new proposed classification methods are used to improve leaf image classification process in [62, 63]. The KNN classifier is used in plant identification and recognition using plant images as in [64-66]. In [67] a new strategy of multi-classification system of wheat leaves diseases is proposed with recognition accuracy of 87.13% by using the KNN classifier. Zernike moments and morphological features are employed with KNN classifier to study classification of plant leaves with a reasonable level of accuracy [68, 69]. KNN classifier is used in ANFIS (Adaptive Neuro-Fuzzy Inference System) and this system was trained by different leaves of a plant with accuracy ratio is 80% [70]. Authors in [71] used the geometrical feature to identify plant leaves based on KNN classifier with recognition rate is 93.17%. Authors in [72] used KNN classifier to propose a new approach to classify the plant leaves by using texture features. In order to improve the performance of plant leaf identification, the Fuzzy KNN classifier is introduced in [73].

3. Probabilistic Neural Network Classifier (PNN)

PNN is considered as a category of ANN which is based on the statistical principle. Classification can be performed by localized networks such as PNN. PNN classifier was proposed by Specht in 1990 [74]. PNN was based on the computing of Bayes statistics which consists of four layers input layer, pattern layer, summation layer and decision layer [75]. PNN classifier is used to classify leaf plants and was showed a high accuracy as in [76, 77]. Shape feature is applied to discuss the plant identification leaves with PNN classifier as in [78] with accepted recognition rate more than 85.6%. In addition, PNN classifier is used in plant leaves classification and recognition based on the center distance for leaf identification and the recognition rate was about 80% [79]. Moreover, PNN classifier is used in different areas of plant classification and recognition as in [80, 81]. A huge number of plant images from different plant species was taken from Flavia data set is trained by using PNN. The methods which are based on half leaf features showed good results when using PNN, and the results showed high recognition accuracy as in [82, 83].

4. Hyper Sphere Classifier
Hyper sphere classifier has the ability to compress the sample data to reduce the space of storage. The main idea of using hyper sphere classifier is to use a hyper sphere represent the points of the cluster. In order to reduce the high dimension of the space, we can use a series of hyper spheres. The overall idea with hyper sphere classifier is based on the approximation of the number of hyper spheres for each sample with the extension of the radius of the hyper sphere and eventually we can use multi sample space and all the sample points. A new classification method based on hyper sphere classifier is presented with accepted level of accuracy in [84].

5. Back Propagation (BP) Classifier

BP is a supervised learning classifier which can compute some functional relationships between its input and output. BP is a multilayer network, fully connected, feedforward network. The first layer is represented by the input while the second layer is represented by the output layer. The hidden layers are defined to be the layers between the input and output layers. Authors in [85] used ANN in the classification process to identify the apple orchard pests. The resulting neural model has been trained by the use of the BP method. A new proposed method to classify the images of plant based on the BP classifier is presented in [86] with a classification rate of 92.7%. The BP is also used to classify the leaves of the plant [87], and the experimental results showed high accuracy level.

6. Random forest (RF) Classifier

The main idea of RF is to build a forest with a random way. The forest consists of decision-making trees. The tasks that are provided by the human experts are useful in Machine Learning applications by using the knowledge about these tasks in the plant classification process as the in case of the veins leaf which need the knowledge of experts [88]. Authors in [89] applied RF algorithm in classification process for LC–MS species of plant species identification. The experimental results showed that the RF is a robust classifier with noisy problems with handling large datasets in fields such as computational biology. A new system of plant identification based on RF is used in [90] with accuracy rate is 88.82%. Moreover, the RF was employed to plant recognition in a complex background in [91] and the accuracy was up to 97.3%. Recently, DNA barcoding as in [92, 93] is used in taxonomic research applications. DNA sequence is used to identify of species to which a plant, animal or fungus are belonging.

7. k-Means Classifier

The k-means clustering algorithm is used as iterative process to divide the data set into another classes in order make the evaluation of clustering technique to provide an optimal performance. Therefore, it represents the mean of all data in the clustering process. However, this classifier is not adequate to process a discrete data, on the other hand it has good results with continuous data. Authors in [94] applied k-means classifier in order to identify the plant diseases but the experimental results showed low performance. A new method for plant identification based on using the k-Means classifier is used is used in the literature with an accepted classification rate.

8. Hybrid Classifiers

In Machine Learning and Data Mining approaches, hybrid classifiers can integrate the results from multiple algorithms with the aim of improving classification accuracy and algorithms performance. There are some commonly used hybrid classifiers that shows good performance and results with the continuous development of plant identification. In addition, plant genes can be employed as alternative mechanism to explore the important plant functions such as the response to the stress. Therefore, authors in [95], proposed a novel computational method for the identification of plant based on hybrid feature system based on combing the (PWM) position weight matrix with the (ID) increment of diversity. In this system, SVM was used to classify constitutive and alternative splice sites. Numerical taxonomy is considered as a method of data mining in the field of botany. In [96], a particle swarm optimization-aided fuzzy cloud classifier is used based on attribute similarity for plant taxonomy. Plant leaves are classified in [97] by using Linear Discriminant Classifier (LDC).

V. Current Limitations and Challenges based Image Processing Techniques

The main objective of this section is to highlight the most crucial limitations and challenges in plant diseases identification to evaluate the performance of the current proposed methods techniques in the literature based on image processing approach.

A. Current problems

Although a lot of methods of plant recognition and diseases detection based on the image were proposed in the last few years, and some of them have achieved recognition results, some problems still exist and summarized in the following.

In the literature review, there are some research studies of the automatic plant identification and its diseases based on image processing techniques. From the analysis process of the current status, there is a gap between the current capabilities of image processing techniques for automatic plant disease identification and the real-world needs. Moreover, most of the used image processing techniques are need more investigations and research to deal with a
wide variety of plant species and diseases identifications of plants. The main challenges and problems in applying image-based methods for plant and its diseases identification can be grouped in the images capturing conditions, the similar symptoms of different diseases, and the noise in the background of images which represent a critical issue and have an impact on performance of the used and applied techniques and methods for plant and its diseases identification. There are some other challenges that affect automatic disease identification that cannot be categorized together with those already discussed such as real-time operation and differences between the distributions of the training data used to learn the model and the data on which the model is to be applied. More researches are needed to focus on new methods for disease identification, for example, based on color transformations, color histograms, pairwise-based classification system, and clustering algorithms. When we propose new features for feature selection method, we should pay attention to whether it is easy or not to extract these features. With the change of time and the impact of the climate change, the shape or color of the pieces of the plant maybe changes even for the same plant. Therefore, this problem should be taken into consideration in the future studies.

The existing classifiers are not designed especially for plant diseases identification. Therefore using such classifiers in classification process of plant diseases recognition, they show low accuracy and performance rates. The researchers are using different data sets, and the size of the data sample is also different in each dataset. It is difficult to evaluate the performance of different algorithms or the classifier.

B. Limitations and challenges

In the next, we provide an analysis of each one of the proposed challenges, emphasizing both the problems that they may cause and how they may have potentially affected the techniques proposed in the literature. The list of these challenges is described below.

1. **Image background.** The segmentation of the ROI (region of interest) for the symptoms of images is usually difficult because of the elements that exist in the background. For instance, sometimes the background of leaves images is noisy with plants or soil or any other elements, then, to segment the ROI this will be considered as a challenge. Therefore, in the last years, the segmentation process of the leaves of plants from a busy background gained a great attention. For instance, Figure 4 shows how the complex background in the image represents a critical challenge.

![Fig. 4. Complex image background: (a) and (b) are multi leaves of tomato plant images](image)

2. **Image capture conditions.** The conditions of capturing plant images are sometimes uncontrolled which can cause difficulty in predicting the characteristics of such images. Also, in practice images have to be captured in the same conditions. Illumination issues are important, where issues like the position of the leaf concerning the sun and overcast conditions. Some of these problems are shown in Figure 5.

![Fig. 5. Uncontrolled environment: (a) and (b) are tomatoes leaves images under uncontrolled conditions.](image)
3. **Feature extraction.** To propose a new feature, we have to take into account whether it is easy to extract or not. With the climate and time change, the shape or color of the plant may be changed. In Figure 6, we can notice that the curve on the leaf is a disease. Moreover, there is another disease on leaf itself which makes the feature extraction is difficult.

![Fig. 6. Tomato leaves image with feature extraction problem](image)

4. **Classifiers and algorithms performance.** When researchers use datasets, they usually use different size of the data sample from these datasets. Therefore, there is difficulty in evaluating the efficiency of the used classifiers and algorithms for plant recognition.

5. **Symptom segmentation.** In some cases of diseases, the symptoms on images do not have defined edges or boundaries which make it difficult to identify. It is important to notice that any change in the boundaries of the ROI may impact on the extracted features to describe those regions. The problem of symptom segmentation lacks the solutions and need more investigations and research in the future. The symptom problems are described in Figure 7.

![Fig. 7. Tomato leaves image with undefined images](image)

6. **Diseases characteristics.** A disease may possess very different characteristics and where it exists on the images of the plant.

7. **Symptom variations.** Different diseases propose different symptoms that may be combined with a hybrid symptom or visually similar that can be difficult to identify among them. Figure 8 depicts the similarity of symptoms of infected leaves.

![Fig. 8. Similar symptoms: (a) and (b) are infected tomato leaves images with many diseases](image)

8. **Similar symptoms with different disorders.** The Similarity of symptoms with different disorders is considered as one of the most important problems in the automatic plant disease diagnosis process.

9. **Training data and model data.** In the automatic plant identification and the plant diseases, there are differences between the distributions of the used data in the training phase than that are used to the data on which the model is to be applied.

10. **Datasets availability.** There is few existing databases are used in the implementation phase of plant identification. Moreover, a kind of plant collected in the available datasets is less and not representative for
deep processing. Moreover, researchers try to collect their datasets themselves which usually do not have a unified rule. As a result, standard benchmark data sets are still challenges.

11. **Fully automated systems.** In some situations, we cannot fully deal with automatic systems that are based on image processing and computer vision methods and algorithms. Plant agricultural engineers and pathologists often have to resort to resources to obtain a reasonable diagnosis of plant diseases. Therefore, diagnosis system should include various modules which have the capability of providing more information about the current problems. This situation can lead a loss of full automation process. However, additional information will be available for a reliable diagnosis.

12. **Real-time Monitoring.** Monitoring the plant in the real time will provide many efforts and ease the diagnosis process. Some applications require real time monitoring, while all the proposed systems yet lack the monitoring in the real time. In the future, monitoring is expected to be a proposing solution of the plant diseases identification.

The use of digital image processing in agriculture is quickly becoming ubiquitous, as emulating human visual capabilities is a fundamental step towards the automation of processes. Creating a computer vision system to perform disease diagnosis and severity measurement is one of the most challenging tasks currently underway [98].

**VI. Conclusions and Future Work**

Plant diseases cause a major impact on the economics of countries in the industrial agricultural worldwide. Nowadays, computer-based methods of plant identification and plant diseases diagnosis is a new research area, with some problems that still need to be solved. Plant science specialists and engineers can depend on the automated diagnosis systems to identify the plant diseases and recognition. The use of digital image processing and computer vision in plant diseases identification and recognition has shown a great importance in identifying plant diseases in the last years, however there are many limitations and challenges that are described in this paper that need to be overcome by solving and minimizing some of the problems that have been mentioned in this research. Also, with the availability of the advances in image processing, strategies, and algorithms that are applied in the last years. With the developments in image processing techniques which can provide a low cost and propose new advancements in plant diseases identification and recognition. As a result, using digital images through image processing area will be more effective representation method in this problem. This paper reviews the state-of-art of the used techniques, methods, and classifiers that are currently used in the literature for plant diseases identification and recognition. Also, limitations, challenges, and future trends are presented in this area.

To overcome some of the proposed limitations shortly in plant recognition and diseases identification field we need to place constraints on capture images condition. However, this strategy needs more additional efforts which are needed to meet those constraints. In the future, machine learning and computer vision techniques such as Markov Random Fields, Graph Theory, Mean Shift, Deep Learning, and Large Margin Nearest Neighbor classification have to be properly explored and used. Moreover, hybrid system techniques will play an important role in plant recognition and diseases identification. Also, plant specialists and experts and plant sciences will depend on the automatic image-based systems to support the decision making of the expertise of human.

Nowadays, scouting is the used method for plant monitoring the stress in trees, for example, but it is an expensive method, time-consuming process, and labor-intensive. In the near future, thanks to the advances in technology as Sensor Networks (SNs), Internet of Things (IoT), and Cloud Computing and their integration this can meet real-time requirements of monitoring would be expected to be easier and available with time. Therefore, there is a necessity to develop reliable plant monitoring systems based on sensors, RFID, etc. that would facilitate advancements and assist in monitoring health of plants and plants identifications and recognition.

**Acknowledgement**

The author would like to acknowledge the valuable comments and suggestions of reviewers, which have improved the quality of the paper.

**References**


[58] A. Hudaya Muhamad Amin, Asad I. Khan, One-Shot Classification of 2-D Leaf Shapes using Distributed Hierarchical Graph Neuron (DHGN) Scheme with k-NN Classifier, Procedia Computer Science 24, 84 – 96, (2013).


